

Imperial College of Science, Technology and Medicine
Department of Physics

**Search for long-lived heavy neutral leptons with
the CMS experiment using machine learning
techniques**

Julia-Suzana Dancu

CERN-THESIS-2023-248
18/07/2023



A dissertation submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy from Imperial College London, October 2023

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

This Thesis presents the analysis strategies and outcomes of a search looking for long-lived Heavy Neutral Leptons (HNLs) in proton-proton collision data corresponding to 137fb^{-1} integrated luminosity, collected by the CMS experiment at the Large Hadron Collider with $\sqrt{s} = 13$ TeV. The HNL particles are searched for via charged-current production and semi-leptonic decay mode into a lepton and jet that can be displaced from the primary collision point. A fine-grained event categorisation scheme ensures enhanced sensitivity to a number of HNL benchmark models including both Dirac- and Majorana-type particles. The search employs a novel deep neural network-based displaced jet tagger and a complementary event-level boosted decision tree algorithm with the aim of separating HNL signal events from Standard Model (SM) background processes. The blinded analysis uses a data-driven background estimation method to evaluate the expected yields, the method being validated in two control regions. After unblinding, no excess of events was found and upper limits on the HNL production cross-section were set for a broad range of HNL mass, lifetime and coupling scenarios, including universal mixing with all SM lepton generations. A particular focus of this Thesis was the evaluation of displaced lepton reconstruction and identification of scale factors, which were found to be one of the dominant systematic uncertainties associated with the signal interpretation. The results presented in this Thesis are competitive with previously published similar searches in high energy physics and are the first in the CMS collaboration to consider arbitrary coupling combinations involving all three SM lepton generations.

Statement of Originality

I, the author of this Thesis, hereby declare the work contained in this document to be my own. All figures labelled “CMS” are sourced directly from CMS publications, including those produced by the author, and referenced in the figure caption. All figures and studies taken from external sources are referenced appropriately throughout this document.

In Chapter 6 I carried out the track-to-vertex association studies and machine learning algorithm development in the context of the end-to-end neural network algorithm for primary vertex reconstruction development efforts for the Phase-2 upgrade of the CMS Level-1 trigger system.

In Chapter 7 I was involved in the development of the analysis strategy, particularly in charge of producing data-to-simulation plots throughout its progress. Furthermore, I developed the event-level binary classifier algorithm (Chapter 8) employed in the data-driven background estimation method used in the analysis, which was developed in collaboration with Vilius Čepaitis. In addition, in Chapter 9 I derived the efficiencies and scale factors applied to the displaced lepton identification and displaced muon reconstruction, treated as systematic uncertainties in the analysis. The tag-and-probe tool used in this study was provided by the CMS collaboration, while the resonance peak fitting algorithm was developed by Mikael Mieskolainen. Finally, the results presented in Chapter 10 were derived by myself, using tools and statistical interpretation strategies developed by the CMS collaboration and colleagues in the analysis team.

The analysis is currently undergoing internal review by the CMS collaboration and is expected to be published soon. The analysis presented in the paper uses a different background estimation method to that presented in this Thesis, however, the results are very compatible, and the same conclusions may be drawn.

Julia-Suzana Dancu

To my beloved grandmother, Kati Nagy
and my always supportive parents.

Acknowledgements

The past five years have been a truly special, yet challenging journey with many ups and downs along the way. I feel extremely privileged and grateful for the opportunity to pursue a PhD in the HEP group at Imperial College London with numerous bright minds, unique talents and inspiring colleagues, sharing their vast knowledge and insightful scientific advice. I would like to thank STFC for their generous scholarship and support for sending me on a long-term attachment to CERN, which has been a fantastic opportunity to get exposed to the core of the HEP scientific community, with many lifetime memories and experiences along the way.

First of all, I would like to thank Alex Tapper for his academic support and leadership through my PhD, his guidance, patience, regular feedback and advice made a huge difference to the outcome of this Thesis.

Secondly, I would like to express my deep gratitude to Vilius Čepaitis for his endless patience, kindness, never-ending help, academic and technical support throughout this PhD. His insightful assistance with the HNL analysis, guidance to successfully complete my tasks, moral and emotional support through the difficulties of the pandemic and PhD project, regular reviews and feedback on my work, his kind encouragement to never give up made this Thesis possible. I feel extremely lucky and blessed to have been given the chance to collaborate and get to know you better. I wish you a bright, successful and fruitful career and to always stay as ambitious and motivated!

In addition, I would like to thank Mikael Mieskolainen and Antoni Shtiplieski for their academic, technical and collegial support, as well as their mentorship and friendship through my PhD studies. Our numerous discussions and debates have been truly inspiring and educational, I am very grateful to have met you and worked with you!

I would like to thank the Imperial College HEP group for their generous support and inspiring research environment where I could grow as a scientist. I would like to thank Gavin Davies for his leadership and support to participate in the SLAC summer school, which has been a unique life experience. Moreover, I would like to thank Paula Brown and the rest of the fundamentals for their administrative support and warm friendship which always brightened up my day. I would also like to thank my fellow IC and CMS colleagues for their friendship and great chats, it has been a privilege to meet such special people: Vukašin Milošević, Sergey Shirobokov, Simone Bologna, Shameena Bonomaly, Răzvan Moise, Ryan Newcombe, Matthew Citron, Edward Scott, Georgios Karathanasis, Kai Hong Law, Giacomo Zecchinelli,

Liv Våge, Christopher Brown, Charlie Naseby, Charlotte Knight and many others. I acknowledge my supervisor Oliver Buchmüller.

I would like to thank the friendship and many great memories shared with amazing people I met throughout my LTA: Kathryn C., Biljana M., Lakshan R., Adrian A. F., Robert A., Daniel P., Ilinca R., Francisco, Ștefana A., Mihaela M. and many, many others. Furthermore, I would like to thank all my wonderful friends in Geneva who helped me get integrated and made my time truly special here: Emanuele S., Sarah P., Alisa and Youri H., Dinah L., Hillel N., Barbara M., Moïse C., Deborah and many others from the local community.

I would like to thank my closest friends for always being there for me throughout my life, for the shared laughs and cries, for the great adventures, memories, chats and opportunity to grow together: Rivky, Liana, Krisztina, Sandy and Jake. Thank you for being such wonderful people through various points in my life!

I would like to thank the long list of mentors, lecturers, teachers and extended family who shaped me through my life and made me become who I am. Your hard work, encouragement, belief and ambition led to the success of me getting this far. I would like to highlight two key people to whom I cannot express my gratitude enough: Ványi Emese and Boga Katalin. You initiated me into the intricacies of mathematics and physics, believed in me from a very young age and guided me towards the path of choosing science as a career. You deserve all the respect and recognition for the often unforgiving and hard work you have been doing as truly outstanding teachers who prepared many generations of students.

Last but not least, I would like to express my eternal gratitude to those who truly made me who I am, those who believed in me from day one, made endless sacrifices and raised me with their unconditional love and care: my parents and my dear grandmother, Kati Nagy. Words cannot express what I feel in my heart towards you, beyond gratitude, love, respect, admiration, blessing. All I can say: thank you!

Contents

1	Introduction	2
2	Theoretical background	7
2.1	The Standard Model of Particle Physics	7
2.1.1	Overview	7
2.1.2	Gauge theories	10
2.1.3	Electroweak symmetry breaking and fermion mass generation	20
2.2	Physics beyond the Standard Model	24
2.2.1	Introduction	24
2.2.2	The Seesaw mechanism	33
2.3	Long-lived particles	40
2.3.1	Introduction	40
2.3.2	Long-lived Heavy Neutral Leptons	43
3	Experimental Setup	47
3.1	The Large Hadron Collider	47
3.2	The CMS detector	52
3.2.1	Tracking system	54
3.2.2	Electromagnetic calorimeter	57
3.2.3	Hadronic calorimeter	59
3.2.4	Solenoid magnet	61
3.2.5	Muon system	61
3.2.6	Data acquisition system	63
3.3	The High-Luminosity LHC and CMS Phase-2 Upgrade	65
3.3.1	Overview	65
3.3.2	Phase-2 Upgrade to the Level-1 trigger system	66

4	Event Reconstruction	69
4.1	Particle Flow reconstruction	69
4.1.1	Track reconstruction	70
4.1.2	Vertex reconstruction	71
4.1.3	Calorimeter cluster reconstruction	71
4.1.4	Muon tracking	72
4.2	Physics object reconstruction	72
4.2.1	Muon reconstruction	72
4.2.2	Electron reconstruction	75
4.2.3	Jet reconstruction	77
4.2.4	Missing transverse energy reconstruction	80
5	Analysis techniques	83
5.1	Monte Carlo event generation	83
5.2	Machine learning	86
5.2.1	Boosted decision trees	87
5.2.2	Neural Networks	90
5.2.3	Domain adaptation	93
5.3	Statistical methods	94
5.3.1	The Pearson correlation coefficient	95
5.3.2	The mutual information coefficient	95
5.3.3	Statistical methods employed to estimate uncertainties	96
5.3.4	Likelihood model	98
5.3.5	Profile likelihood ratio	99
5.3.6	The ABCD method	100
5.3.7	Hypothesis testing	102
6	Track-to-vertex association for the Phase-2 upgrade of the Level-1 trigger system	108
6.1	Introduction	108
6.1.1	Baseline vertex reconstruction and track association algorithms	109
6.2	Training samples	110
6.3	Track-to-vertex association with MVA methods	113
6.3.1	Association using boosted decision trees	113
6.3.2	Association using dense and convolutional neural networks	115

6.3.3	Performance of the MVA models considered for track-to-vertex association	116
6.4	End-to-end neural network for vertexing	118
6.4.1	Epilogue	119
6.5	Summary	120
7	Search for long-lived heavy neutral leptons with the CMS detector	122
7.1	Heavy neutral lepton signatures	123
7.1.1	Heavy neutral lepton signal simulation	125
7.2	Background processes	130
7.3	Trigger strategy and data sets	132
7.4	Event selection and categorisation	133
7.4.1	Physics objects	134
7.4.2	Preselection of events	139
7.4.3	Analysis selections	139
7.4.4	Event categorisation	146
7.4.5	Corrections to data and simulation	147
7.5	Deep neural network-based displaced jet tagger for heavy neutral lepton searches	148
7.5.1	Displaced signal jet subclasses	149
7.5.2	Standard model background jet classes	149
7.5.3	Deep neural network architecture	150
7.5.4	Tagger performance and validation	151
7.6	Summary	152
8	Event-level classifier for the long-lived heavy neutral lepton search	157
8.1	Introduction	157
8.2	Correlation studies	160
8.3	Data resampling	166
8.4	The boosted decision tree model evaluation and validation	168
8.5	Summary	173

9 Displaced lepton systematic uncertainty studies for the long-lived heavy neutral lepton search	181
9.1 Sources of systematic uncertainty in the search for long-lived heavy neutral leptons	181
9.1.1 Signal systematic uncertainties	182
9.1.2 Background systematic uncertainties	183
9.2 Displaced muon efficiencies	183
9.2.1 Displaced muon identification efficiency	183
9.2.2 Displaced muon track reconstruction efficiency	191
9.3 Displaced electron efficiencies	197
9.3.1 Displaced electron identification scale factors	197
9.4 Summary	214
10 Background estimation and results	216
10.1 Background estimation	216
10.1.1 Introduction	216
10.1.2 Threshold optimisation	217
10.1.3 Closure tests	218
10.2 Results	221
10.2.1 Limits in mass-coupling plane	221
10.2.2 Discussion of the results	223
10.3 Summary	225
11 Conclusions	228
A Simple error propagation equation	248
B Heatmaps for the track-to-vertex association studies	249
C Control region validation plots	251
D Complementary material to the displaced muon scale factor studies	278
D.1 Displaced muon ID scale factors	278
D.2 Displaced muon reconstruction scale factors	278

E	Complementary material to the displaced electron ID scale factor studies	282
E.1	Data to MC agreement plots in the 3-lepton CR	282
E.2	Displaced electron ID efficiencies	288
E.3	Displaced electron ID scale factors	293
E.4	Displaced electron ID scale factor systematic uncertainty study	303
F	Complementary material to the BDT event classifier study	309
F.1	2D heatmaps between the tagger score and BDT input features . . .	309
F.2	2D heatmaps between the $m(\ell_1, \ell_2, j^*)$ variable and BDT input features	314
F.3	Correlation studies between the $m(\ell_1, \ell_2, j^*)$ variable and BDT inputs	318
F.4	Performance plots of the correlated BDT model	321
F.5	Complete nuisance parameter pulls and impacts	324

List of Tables

2.1	Summary of the twelve fundamental fermions grouped by leptons and quarks, as well as by generation [24].	9
2.2	Summary of SM fundamental forces with their associated mediator spin-1 gauge bosons, indicating the gauge symmetries that govern them. Since the Higgs boson is a spin-0 scalar boson, it is not a force carrier particle.	10
2.3	Table presenting the forces experienced by fermions.	10
4.1	The applied custom loose cut-based electron identification criteria. . .	77
4.2	Jet identification criteria for 2016 and 2017/2018 data-taking scenarios.	79
7.1	Branching ratios calculated by MADGRAPH for the tree-level decay modes of Dirac HNLs at various masses.	127
7.2	p_T thresholds employed by the single lepton triggers used in the analysis for all three data-taking years. Offline p_T thresholds were increased by 2 GeV to improve the selection efficiency.	132
7.3	Overview of the preselection requirements, as well as definitions of the signal and control regions for the analysis. The kinematic selection criteria for tight leptons are shown per data-taking year (2016-2018).	140
8.1	SM MC background samples used for training the BDT event classifier, in each data-taking year, respectively. The table presents the number of events ($\#$ evts), the fraction of positive-weighted events (f_{pos}) and the corresponding cross-sections used for normalisation (σ).	159
8.2	Predicted relative ratios between background MC processes per data-taking year used in order to subsample the background simulation sets for the BDT training.	168
8.3	Input variables used in the two BDT models considered. The first BDT model considered is uncorrelated with the displaced jet tagger and $m(\ell_1, \ell_2, j^*)$ variable, whereas the second BDT model is highly dependent on them.	171

9.1	Table presenting the corresponding luminosities for each HLT path in data per data-taking year.	185
9.2	Scale factor values evaluated for displaced muon loose identification criteria as a function of the muon p_T and $ d_{xy}^{sig} $ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events for all three data-taking years. These values have been implemented in the analysis framework as correction scale factors.	189
9.3	Scale factor values evaluated for displaced muon loose identification criteria as a function of the muon p_T and $ d_{xy} $ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events.	190
9.4	Scale factor values evaluated for displaced muon reconstruction as a function of the muon p_T and $ d_{xy} $ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events for all three data-taking years. These scale factors are eventually propagated in the analysis framework as corrections.	196
9.5	Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data-taking years.	209
9.6	Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}(e)$ kinematic variable pair for all three data-taking years.	210
9.7	Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data-taking years. Uncertainties presented are statistical only.	211
9.8	Pull values estimated between the “raw” and “geometric mean” scale factor approaches on the $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data-taking years.	212
9.9	Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data-taking years including systematic uncertainties. These scale factors are eventually propagated further in the HNL analysis. .	213
10.1	Effect of systematic uncertainties on the HNL signal cross-section expressed as percentage impacts. The impacts on the background yields in the sideband regions are averaged over all 48 categories of the analysis. The bottom five uncertainties affect the background estimation yield, whereas the rest of the entries affect the HNL signal yields.	226

List of Figures

2.1	Fermion masses structured by their hierarchical order in 3 generations. The neutrinos are denoted as ν_i by their mass eigenstates and normal ordering is assumed. Since their exact masses are unknown, mass ranges are shown based on neutrino oscillation experiments. Figure taken from Ref. [38].	28
2.2	Exclusion limits on the mixing between the electron-, muon- and tau-neutrino and a single heavy neutrino in the mass range of 10 MeV – 100 GeV, as well as the expected sensitivity from future experiments [59].	35
2.3	Feynman diagrams presenting various HNL production channels via mixing with active neutrinos of flavour $\alpha = e, \mu, \tau$. The left and middle diagrams show HNL production via EW gauge boson on-shell decays, while the diagram on the right shows HNL production via charged meson decay.	38
2.4	Feynman diagrams representing the leptonic and semi-leptonic HNL decay channels via mixing with active neutrinos of flavour $\beta = e, \mu, \tau$. The left diagram shows an HNL decaying via an off-shell W^* boson, while the diagram on the right shows an HNL decay via an off-shell Z^* boson. γ marks the flavour of the final state leptons resulting from the off-shell bosons.	39
2.5	The range of normalised flavour mixings between HNLs and active neutrinos, consistent with the current neutrino oscillation data, assuming degenerate heavier sterile neutrino states. The different contours correspond to the allowed case of normal hierarchy (red) and inverted hierarchy (blue) within the shown uncertainties. Figure taken from Ref. [59].	40
2.6	Particle masses and lifetimes for both elementary and composite SM particles. Prompt and stable particles are marked by the shaded regions, while lifetimes of 10 μm up to 10 m are associated with LLPs decaying inside a general-purpose detector. Figure from Ref. [65]. . .	42
2.7	Schematic view of a variety of BSM LLP signatures in general-purpose detectors at the LHC shown in their transverse plane. Figure from Ref. [66].	43

2.8	Contour lines for various HNL decay lengths, $c\tau_0$, assuming equal coupling to all three neutrino generations simultaneously. Shaded areas correspond to approximate current exclusion limits by different experiments. Figure from Ref. [67].	44
3.1	Schematic diagram of the CERN accelerator complex, highlighting various accelerator stages before the proton beams are injected into the LHC superconducting synchrotron ring. In addition, other particle extraction facilities and experiments are shown. Image credit: CERN.	50
3.2	Diagrams showing the cumulative integrated luminosity versus day (left) and over the total operation period (right) recorded by CMS for p-p collisions at nominal centre-of-mass energy. The right diagram also shows the total integrated luminosity delivered by the LHC [74].	51
3.3	Distribution of the average number of interactions per bunch crossing (pileup) for p-p collisions during Run 2 (13 TeV) operation of the CMS detector. The overall mean values and the minimum bias cross-sections are shown [74].	51
3.4	Exploded views of the CMS detector showing an onion-like layer structure and the effects of the strong magnetic field (2 and 3.8 T, respectively) on charged particles (diagrams from [77, 78]). The left figure is a section of the detector in its transverse view. The right figure is a 3-dimensional projection of CMS enhancing the concentric nature of the sub-detector layers (tracker, electromagnetic and hadronic calorimeters, superconducting solenoid and muon chambers, as well as the forward hadronic calorimeter) around the collision point in the centre of the beamline.	52
3.5	Schematic diagram of one-quarter of the Phase-1 CMS tracking system in r-z view. The pixel detector is closest to the interaction point and is shown in green. The silicon strip detectors are depicted as red and blue segments depending on whether they are single-sided or double-sided strip modules, respectively. Figure adapted from [81]. . .	55
3.6	Schematic diagram of one-quarter of the CMS ECAL in r-z view, presenting the layout of the lead tungstate crystals in the electromagnetic barrel (EB) and endcap detectors (EE), as well as the position of the pre-shower detector in front of the EE. Figure taken from [84]. . . .	58
3.7	Schematic diagram of one-quarter of the CMS HCAL in r-z view, presenting the layout of the sampling detectors in the hadronic barrel (HB), endcap detectors (HE) and outer (HO) detectors, as well as the position of the forward hadronic (HF) detector down the beamline. Figure taken from [75].	59

3.8	Schematic diagram of one-quarter of the CMS detector in r-z view, presenting the layout of all detector subsystems, but with a focus on the structure of the muon system. The various muon chambers are interleaved between the iron flux return yoke. The drift tubes (DTs) are located in the barrel region (labelled as "MB" in orange), while the cathode strip chambers (CSCs) are located in the endcaps of the muon system (labelled "ME" in green). The resistive plate chambers (RPCs) are located in both the barrel and endcap regions of the muon system (labelled "RB" and "RE" in blue). Figure taken from [90].	62
3.9	Schematic view of the upgraded Phase-I Level-1 trigger architecture used during the Run-2 data collection period of the CMS experiment [96].	65
3.10	Functional diagram of the components to the Phase-2 Level-1 Trigger architecture [99].	67
4.1	Flow-chart describing the different types of muons reconstructed within the PF algorithm in CMS. Segments are made of several hits. Global muons consist of standalone muon tracks reconstructed in the muon spectrometer and then extrapolated inwards to the tracker, finding compatible tracks. Tracker muons ensure that a track reconstructed in the tracker has matching hits in the muon station.	73
4.2	Signal muons from HNL decay reconstruction and identification efficiencies as a function of generator-matched lab-frame p_T and transverse displacement. Reconstruction efficiencies show a drop at low- p_T and with increasing L_{xy} . Figures from [109].	75
4.3	Signal electrons from HNL decay reconstruction and identification efficiencies as a function of generator-matched lab-frame p_T and transverse displacement. Electron reconstruction efficiencies show poorer performance compared to muons due to the more complex signatures they leave in the CMS detector system. Figures from [109].	77
4.4	Signal jets from HNL decay reconstruction and identification efficiencies as a function of generator-matched lab-frame p_T and transverse displacement. Figures from [109].	79
5.1	Schematic diagram of a simple binary decision tree where S is signal while B is background.	88
5.2	The cross-entropy validation function development the training and testing data samples as the number of BDT training iterations is increased. There is an optimal cut for the iteration number to maximise training on the data sample while minimising overtraining.	90
5.3	Schematic sketch of a multilayer perceptron or dense neural network with one input layer, two hidden layers and an output layer of two nodes corresponding to a binary classifier model.	92

5.4	Schematic diagram of a NN employing domain adaptation by back-propagation via introducing a gradient reversal layer (bottom pink branch). The NN has a feed-forward architecture with a feature extractor (green) and class label predictor (blue) branch. The DA branch connects to the feature extractor branch and has the role of domain label classifier [134].	94
5.5	Schematic diagram of the ABCD background estimation technique definition where D is the blinded signal region of the analysis. . . .	101
6.1	Event display of the FastHisto clustering algorithm, illustrating its vertex association in the z_0 position along the beamline. The true vertex in the event is contained within a fixed window since its associated tracks are reconstructed with a resolution. The left narrow peak is a "fake" track that was incorrectly constructed by the Kalman Filter by accidentally matching unrelated hits in the tracker. Its narrowness is a hint of its unphysical nature.	110
6.2	Track and event feature distributions of selected signal and PU tracks.	111
6.3	Quantifying feature correlations via scatter plots for $t\bar{t}$ ("true") and PU tracks ("false").	112
6.4	Graphs presenting the possible correlations between the AUC and the logarithmic loss for the various BDT models trained on the $t\bar{t}$ sample. In the left plot, the number of input features for individual models is illustrated in a different colour while the right plot shows the impact on the BDT performance of the inclusion of specific input variables, such as δz and p_T	114
6.5	Quantifying relative feature importances in the BDT model for $t\bar{t}$ (left) and $Z \rightarrow \mu\mu$ (right) signal processes.	115
6.6	Classification accuracy (left) and cross-entropy loss (middle) development by training epoch of the dense NN model, as well as in the test data set. The dense NN output (right) shows good separation in the test set. Top plots show the dNN training and test performance for $t\bar{t}$ events, whereas the bottom was trained on $Z \rightarrow \mu\mu$ events.	116
6.7	Classification accuracy (left) and cross-entropy loss (middle) development by training epoch of the convoluted NN model, as well as in the test data set. The CNN output (right) shows good separation in the test set. The top plots show the CNN training and test performance for $t\bar{t}$ events, whereas the bottom was trained on $Z \rightarrow \mu\mu$ events. . .	117
6.8	ROC curves (top) and purity vs matching efficiency (bottom) curves comparing the performance of the four track-to-vertex association models investigated in this study. The left plots show the training performance on the $t\bar{t}$ data set, whereas the middle ones on $Z \rightarrow \mu\mu$ events. The right plots compare the performance of the two NNs on both data sets.	118

6.9	Schematic of the proposed end-to-end NN architecture for PV reconstruction and track-to-vertex association for the Phase-2 L1T upgrade [99].	119
7.1	Feynman diagrams presenting the CC HNL production and semi-leptonic decay channels considered in this analysis. Charge-conjugate modes of the on- and off-shell bosons are also considered. The left diagram shows an opposite-sign dilepton final state available for both Dirac and Majorana HNL scenarios, whereas the right diagram shows a same-sign dilepton final state possible only for Majorana-type HNLs. The latter is an instance of LNV.	123
7.2	Generated HNL signal samples using MADGRAPH in the mass-coupling plane for both Majorana (left) and Dirac (right) scenarios, with equal couplings to all three lepton generations ($ V_e = V_\mu = V_\tau $). The colour-coded, tilted lines correspond to HNLs with the same proper lifetime.	126
7.3	The 66 relative coupling points used for each HNL signal sample generated with equal couplings to the 3 SM neutrino generations $ V_e = V_\mu = V_\tau $. These coupling points are expressed in barycentric coordinates ($f_e + f_\mu + f_\tau = 1$).	127
7.4	Generator-level kinematic distributions of HNL particles and associated leptons for various benchmark models and coupling scenarios: (a) HNL transverse momentum; (b) HNL pseudorapidity; (c) HNL boost; (d) HNL transverse displacement; (e) ℓ_1 transverse momentum; (f) ℓ_1 pseudorapidity; (g) ℓ_2 transverse momentum; (h) ℓ_2 pseudorapidity [109].	129
7.5	Cutflow diagram presenting the selection of jets in the analysis. . . .	136
7.6	Diagrams presenting the categorisation of events by separation of the jet and the subleading lepton, ℓ_2	136
7.7	Reconstructed event observables for various benchmark HNL scenarios assuming equal coupling to the first two generations, motivating the selection criteria applied in the preselection of the analysis, as well as SR and CRs [109].	141
7.8	Cumulative cutflow diagram presenting event-level preselection and SR cuts after the trigger selection was applied for a representative HNL signal model (Majorana-type of $m_N = 10\text{GeV}$, $c\tau_0 = 1\text{mm}$), the two main SM background processes and data (2016).	142
7.9	Distribution of the HNL candidate jet p_T in the DY control region. The distributions are shown for the 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	143

- 7.10 Distribution of the ℓ_1 transverse momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows. 144
- 7.11 Distribution of the ℓ_2 transverse momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows. 145
- 7.12 Distribution of p_T^{miss} in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows. 146
- 7.13 Schematic diagram of the DNN-based displaced jet tagger architecture with the goal to capture the decay products of long-lived HNLs searched in this analysis. 150
- 7.14 Tagger ROC curves for signal (a) muon-jets; (b) electron-jets and (c) quark-jets versus the equivalent SM background jet class taken from an exclusive W+jets production sample, measured for an HNL with $m_N = 4.5 \text{ GeV}$ with various proper lifetimes. 153
- 7.15 Modelling of the tagger output in the DY control region, resolved muon scenario. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OS dimuon category in the first row, and the SS electron-muon mixed flavour category in the bottom row. The poor agreement in the SS category results from the lack of QCD multijet events, which are one of the main background sources for this category. 154
- 7.16 Modelling of the tagger output in the DY control region, merged muon scenario. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OS dimuon category in the first row, and the SS electron-muon mixed flavour category in the bottom row. 154
- 7.17 Modelling of the tagger output in the DY control region, merged electron scenario. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OS dielectron category in the first row, and the SS muon-electron mixed flavour category in the bottom row. 155
- 8.1 Correlation matrix presenting the linear (anti)correlation strength between all possible BDT input feature pairs. The matrix is based on the signal simulation in the SR for the 2016 data-taking period. . . . 161

8.2	Correlation matrix presenting the linear (anti)correlation strength between all possible BDT input feature pairs. The matrix is based on the background simulation in the SR for the 2016 data-taking period.	163
8.3	Correlation matrix, presenting the higher-order correlation strength between all possible BDT input feature pairs. The matrix is based on the signal simulation in the SR for the 2016 data-taking period.	164
8.4	Correlation matrix, presenting the higher-order correlation strength between all BDT input feature pairs. The matrix is based on the background simulation in the SR for the 2016 data-taking period.	165
8.5	BDT input feature ranking by the extent of correlation with the tagger score, using the Pearson correlation coefficient (2016 simulation).	166
8.6	BDT input feature ranking the by extent of correlation with the tagger score, using the mutual information metric (2016 simulation).	167
8.7	The relative background ratios of MC simulated events in the signal region, for each data-taking year. All dilepton flavour and charge categories are combined for each year. For the BDT training, simulated background processes with LO precision were employed.	169
8.8	Classification error development by training epoch for the uncorrelated BDT model, as well as in the test data set.	170
8.9	Logarithmic loss development by training epoch for the uncorrelated BDT model, as well as in the test data set.	170
8.10	Feature importances weighted by <i>weight</i> for the two BDT models considered (2016 simulation): left plot corresponds to the uncorrelated BDT; right plot corresponds to the correlated BDT.	171
8.11	2D heatmaps showing almost no correlation between the score of the uncorrelated BDT model and displaced jet tagger score (2016 simulated background (top four) and signal (bottom four) processes), categorized by dilepton flavour and combined dilepton charge in the SR of the HNL search.	174
8.12	BDT performance ROC curves on the evaluation data set for the uncorrelated BDT, presented per separate data-taking years.	175
8.13	Variable scan performed on all BDT input features for all three HNL benchmark samples used for evaluation studies, employing the 2016 data set.	176
8.14	BDT score distributions for the uncorrelated model, showing good separation between background and signal processes in most categories (based on 2016 MC simulation).	177

8.15	Distribution of the uncorrelated BDT model score in the DY control region. The distributions are shown for the 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	178
8.16	Distribution of the uncorrelated BDT model score in the resolved control region. The distributions are shown for the 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	179
9.1	Fitted peaks for simulated (top row) and data (bottom row) “pass” and “fail” probe muons in the first bin of the p_T vs $ d_{xy}^{sig} $ parameter space (2016).	186
9.2	Fitted peaks for simulated (top row) and data (bottom row) “pass” and “fail” probe muons in the first bin of the p_T vs $ d_{xy} $ parameter space (2016).	186
9.3	Identification efficiencies of data (left) and MC (right) as a function of the probe muon p_T vs $ d_{xy}^{sig} $ for all 3 data-taking years (2016 to 2018, top to bottom).	187
9.4	Identification efficiencies of data (left) and MC (right) as a function of the probe muon p_T vs $ d_{xy} $ for all 3 data-taking years (2016 to 2018, top to bottom).	188
9.5	Fitted peaks for simulated (top row) and data (bottom row) “pass” and “fail” probe tracker muons and tracks in the first bin of the p_T vs $ d_{xy} $ parameter space (2016).	192
9.6	Fitted peaks for simulated (top row) and data (bottom row) “pass” and “fail” probe tracker muons and tracks in a mid-range bin of the p_T vs $ d_{xy} $ parameter space (2016).	193
9.7	Fitted peaks for simulated (top row) and data (bottom row) “pass” and “fail” probe tracker muons and tracks in the highest bin of the p_T vs $ d_{xy} $ parameter space (2016).	194
9.8	Reconstruction efficiencies of data (left) and MC (right) as a function of the probe muon p_T vs $ d_{xy} $ for all 3 data-taking years (2016 to 2018, top to bottom).	195
9.9	Feynman diagrams of external (real) and internal (virtual) photon conversions into electron-positron pairs.	198

9.10	Generator-matched distributions aiming to select the optimal phase space enriched in genuine photon conversions in order to maximize purity in the control region used in the asymmetric photon conversion study, using simulation from 2016. The top row shows the trilepton mass vs dimuon mass distributions for signal (left), background (middle) and signal purity (right) highlighting the region of optimal purity selection. The bottom row presents the opening angle $\Delta R(\ell_1 e)$ between the leading muon and trailing electron vs ΔR between the subleading muon and trailing electron.	199
9.11	Data to MC agreement plots in the 3-lepton CR for dilepton mass $m(\ell_1 \ell_2)$, trilepton mass $m(\ell_1 \ell_2 e)$ and trilepton p_T . Plots on the left do not have custom electron ID applied whereas those on the right do.	202
9.12	Data to MC agreement plots in the 3-lepton CR for the trailing lepton $d_{xy}(e)$, $d_{xy}^{sig}(e)$ and trilepton charge. Plots on the left do not have custom electron ID applied whereas those on the right do.	203
9.13	Data to MC agreement plots in the 3-lepton CR for the leading, subleading and trailing lepton p_T distributions. Plots on the left do not have custom electron ID applied whereas those on the right do. .	204
9.14	Efficiency and scale factor distributions evaluated for displaced custom electron identification performance for a number of kinematic variables of interest (2016).	205
9.15	1D efficiency and scale factor distributions evaluated for displaced electron custom identification performance for $d_{xy}(e)$, $d_{xy}^{sig}(e)$ and $p_T(e)$ kinematic variables for 2017 (top) and 2018 (bottom) data-taking years.	206
9.16	1D efficiency distributions evaluated for reconstructed signal electrons from 3 HNL benchmark models using custom identification. The distributions are shown for $d_{xy}(e)$ (top), $d_{xy}^{sig}(e)$ (middle) and $p_T(e)$ (bottom) kinematic variables in the three data-taking years 2016, 2017 and 2018 (left to right).	207
9.17	2D efficiency distributions evaluated for displaced electron identification on both data (left) and MC (right) for kinematic variable pairs of interest (2016).	208
10.1	Schematic diagram of the ABCD background estimation technique definition applied for the long-lived HNL search, where D is the blinded SR of the analysis. The low tagger score, P_{low} , region (shaded) is removed to maximise signal-like jets being selected.	218
10.2	Outcome of the closure test performed in the low-mass region VR1 for merged (left) and resolved (right) event categories for all three data-taking years (2016-2018 from top to bottom).	219

10.3	Outcome of the closure test performed in the tagger sideband VR2 for merged (left) and resolved (right) event categories for all three data-taking years (2016-2018 from top to bottom).	220
10.4	The predicted saturated likelihood test statistic distributions generated with toy-data and the observed value in data (blue arrow) for both the VR1 (left) and VR2 (right) GOF tests.	221
10.5	Observed and predicted yields in the SR for merged (left) and resolved (right) event categories for all three data-taking years (2016-2018 from top to bottom). Expected yields for two benchmark HNL models are overlaid.	222
10.6	The predicted saturated likelihood test statistic distribution generated with toy-data and the observed value (blue arrow) for the SR GOF test.	223
10.7	Expected and observed limits for Dirac and Majorana HNL for different benchmark coupling scenarios.	224
B.1	Plots presenting the 2D distribution and lack of correlation between δz and p_T , as well as z_0 and δz for $t\bar{t}$ tracks and PU. There is a tendency towards $\delta z = 0$, particularly for some of the $t\bar{t}$ tracks.	249
B.2	Heatmaps presenting the uniform distributions of the p_T vs z_0 , η vs z_0 , as well as $\tan \lambda$ vs z_0 proving no correlation between them for $t\bar{t}$ tracks and PU.	250
C.1	Distribution of the HNL candidate jet transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows. . .	252
C.2	Distribution of the HNL candidate jet transverse momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.	253
C.3	Distribution of the HNL candidate jet transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows. . .	254
C.4	Distribution of the ℓ_1 transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	255

C.5	Distribution of the ℓ_1 momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.	256
C.6	Distribution of the ℓ_1 transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.	257
C.7	Distribution of the ℓ_2 transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	258
C.8	Distribution of the ℓ_2 momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.	259
C.9	Distribution of the ℓ_2 transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.	260
C.10	Distribution of the p_T^{miss} in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	261
C.11	Distribution of the p_T^{miss} in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.	262
C.12	Distribution of the p_T^{miss} in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.	263

C.13 Distribution of the dilepton mass in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	264
C.14 Distribution of the dilepton mass in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	265
C.15 Distribution of the 3-body invariant mass variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	266
C.16 Distribution of the 3-body invariant mass variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	267
C.17 Distribution of the sphericity event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	268
C.18 Distribution of the sphericity event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	269
C.19 Distribution of the aplanarity event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	270
C.20 Distribution of the aplanarity event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	271

C.21	Distribution of the C-parameter event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	272
C.22	Distribution of the C-parameter event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	273
C.23	Distribution of the isotropy event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	274
C.24	Distribution of the isotropy event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	275
C.25	Distribution of the circularity event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	276
C.26	Distribution of the circularity event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.	277
D.1	Scale factor values evaluated for displaced muon loose identification criteria as a function of the muon p_T and $ d_{xy}^{sig} $ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events. These values have been implemented in the analysis framework as correction scale factors.	279
D.2	Scale factor values evaluated for displaced muon loose identification criteria as a function of the muon p_T and $ d_{xy} $ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events.	280
D.3	Scale factor values evaluated for displaced muon reconstruction as a function of the muon p_T and $ d_{xy} $ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events. These scale factors are eventually propagated in the analysis framework as corrections. . . .	281

E.1	Data to MC agreement plots in the 3-lepton CR for the leading, subleading and trailing lepton η distributions. Plots on the left do not have custom electron ID applied whereas those on the right do.	283
E.2	Data to MC agreement plots in the 3-lepton CR for the leading, subleading and trailing lepton ϕ distributions. Plots on the left do not have custom electron ID applied whereas those on the right do.	284
E.3	Data to MC agreement plots in the 3-lepton CR for $\Delta\eta(\ell_1 e)$ and $\Delta\eta(\ell_2 e)$. Plots on the left do not have custom electron ID applied whereas those on the right do.	285
E.4	Data to MC agreement plots in the 3-lepton CR for $\Delta\phi(\ell_1 e)$ and $\Delta\phi(\ell_2 e)$. Plots on the left do not have custom electron ID applied whereas those on the right do.	286
E.5	Data to MC agreement plots in the 3-lepton CR for $\Delta R(\ell_1 e)$ and $\Delta R(\ell_2 e)$. Plots on the left do not have custom electron ID applied whereas those on the right do.	287
E.6	2D efficiency distributions evaluated for displaced electron identification using the raw method on both data (left) and MC (right) for kinematic variable pairs of interest (2017).	288
E.7	2D efficiency distributions evaluated for displaced electron identification using the raw method on both data (left) and MC (right) for kinematic variable pairs of interest (2018).	289
E.8	2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for kinematic variable pairs of interest (2016).	290
E.9	2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for kinematic variable pairs of interest (2017).	291
E.10	2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for kinematic variable pairs of interest (2018).	292
E.11	Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years.	293
E.12	Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}(e)$ kinematic variable pair for all three data taking years.	294
E.13	Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years. Uncertainties presented are statistical only.	295

E.14	Pull values estimated between the “raw” and “geometric mean” scale factor approaches on the $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years.	296
E.15	Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years including systematic uncertainties. These scale factors are eventually propagated further in the HNL analysis. .	297
E.16	Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $\eta(e)$ kinematic variable pair for all three data taking years.	298
E.17	Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}(e)$ kinematic variable pair for all three data taking years.	299
E.18	Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $\eta(e)$ kinematic variable pair for all three data taking years.	300
E.19	Pull values estimated between the “raw” and “geometric mean” scale factor approaches on the $p_T(e)$ vs $d_{xy}(e)$ kinematic variable pair for all three data taking years.	301
E.20	Pull values estimated between the “raw” and “geometric mean” scale factor approaches on the $p_T(e)$ vs $\eta(e)$ kinematic variable pair for all three data taking years.	302
E.21	2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years in the low pile-up region.	303
E.22	2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years in the high pile-up region.	304
E.23	Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years in the low pile-up region.	305
E.24	Pull values estimated between the inclusive scale factors and those evaluated in the low pile-up region on the $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years.	306
E.25	Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years in the high pile-up region.	307

E.26	Pull values estimated between the inclusive scale factors and those evaluated in the high pile-up region on the $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years.	308
F.1	2D heatmaps between the tagger score and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).	310
F.2	2D heatmaps between the tagger score and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).	311
F.3	2D heatmaps between the tagger score and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).	312
F.4	2D heatmaps between the tagger score and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).	313
F.5	2D heatmaps between $m(\ell_1, \ell_2, j^*)$ and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).	314
F.6	2D heatmaps between $m(\ell_1, \ell_2, j^*)$ and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).	315

F.7	2D heatmaps between $m(\ell_1, \ell_2, j^*)$ and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).	316
F.8	2D heatmaps between $m(\ell_1, \ell_2, j^*)$ and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).	317
F.9	BDT input feature ranking by extent of correlation to $m(\ell_1, \ell_2, j^*)$, using the Pearson correlation coefficient (2016 simulation).	319
F.10	BDT input feature ranking by extent of correlation to $m(\ell_1, \ell_2, j^*)$, using the mutual information metric (2016 simulation).	320
F.11	Classification error development by training epoch of the correlated BDT model, as well as in the test data set.	321
F.12	Logarithmic loss development by training epoch of the correlated BDT model, as well as in the test data set.	321
F.13	BDT performance ROC curves on the evaluation data set for the correlated BDT, trained for all 3 data taking years.	321
F.14	BDT score distributions of the correlated model, showing good separation between background and signal processes (2016 MC simulation).	322
F.15	2D heatmaps between the BDT score of the correlated model and displaced jet tagger information, categorized by dilepton flavour and combined dilepton charge. Stronger correlation between the tagger information and the BDT output is observed for both background (top row) and signal (bottom row) 2016 MC simulation processes.	323
F.16	NP pulls ranked by decreasing impact values. Impacts 1–60 are shown. Signal model used: $m_N = 10 \text{ GeV}$, $c\tau_0 = 1 \text{ mm}$, $V_e = V_\mu = V_\tau$	325
F.17	NP pulls ranked by decreasing impact values. Impacts 61–120 are shown. Signal model used: $m_N = 10 \text{ GeV}$, $c\tau_0 = 1 \text{ mm}$, $V_e = V_\mu = V_\tau$	326
F.18	NP pulls ranked by decreasing impact values. Impacts 121–180 are shown. Signal model used: $m_N = 10 \text{ GeV}$, $c\tau_0 = 1 \text{ mm}$, $V_e = V_\mu = V_\tau$	327
F.19	NP pulls ranked by decreasing impact values. Impacts 181–240 are shown. Signal model used: $m_N = 10 \text{ GeV}$, $c\tau_0 = 1 \text{ mm}$, $V_e = V_\mu = V_\tau$	328
F.20	NP pulls ranked by decreasing impact values. Impacts 241–300 are shown. Signal model used: $m_N = 10 \text{ GeV}$, $c\tau_0 = 1 \text{ mm}$, $V_e = V_\mu = V_\tau$	329

F.21 NP pulls ranked by decreasing impact values. Impacts 301–360 are shown. Signal model used: $m_N = 10 \text{ GeV}$, $c\tau_0 = 1 \text{ mm}$, $V_e = V_\mu = V_\tau$. . 330

F.22 NP pulls ranked by decreasing impact values. Impacts 361–420 are shown. Signal model used: $m_N = 10 \text{ GeV}$, $c\tau_0 = 1 \text{ mm}$, $V_e = V_\mu = V_\tau$. . 331

F.23 NP pulls ranked by decreasing impact values. Impacts 421–480 are shown. Signal model used: $m_N = 10 \text{ GeV}$, $c\tau_0 = 1 \text{ mm}$, $V_e = V_\mu = V_\tau$. . 332

Chapter 1

Introduction

The *Standard Model* (SM) of particle physics is the most comprehensive theory that describes mathematically the nature and interactions of fundamental particles and forces governing all visible matter in the Universe. The SM has been extensively tested experimentally and is known as one of the most successful scientific theories of the twentieth century. Nevertheless, the SM is not the ultimate theory of Nature, as it lacks a description of or justifying some empirical observations made about the Universe. For instance, it only includes three fundamental forces that typically manifest at the quantum level, while it fails to describe the gravitational force that governs matter at macroscopic levels. In addition, the SM does not give any explanation for the existence or nature of *Dark Matter*, despite numerous empirical observations of its effect in the cosmos. Moreover, the Universe mainly consists of matter particles, while predictions of the early Universe postulate that equal amounts of matter and antimatter were produced after the Big Bang. The SM gives no explanation for the extent of *matter-antimatter asymmetry* observed in the current-day Universe.

Experimental tests conducted in the field of *high energy physics* (HEP) reach an impressive TeV energy scale. Large-scale accelerator complexes, such as the *Large Hadron Collider* (LHC), located at the *European Organisation for Nuclear Research* (CERN), achieve the highest particle collision energies mankind has ever produced in a controlled laboratory environment. However, these energy scales are still about fifteen orders of magnitude away from the *Planck scale*, beyond which the laws of physics are expected to break down and quantum gravity effects are expected to dominate. Therefore, the SM might be an *effective theory* of an even more fundamental theory governing higher energy physics processes.

The SM assumes that neutrinos are massless elementary particles governed by the

weak force. The observation of neutrino oscillations, and the mixing between several neutrino flavours in flight, established that at least two of the SM neutrinos have mass. The evidence for non-zero neutrino masses is arguably the first indication for *physics beyond the Standard Model* (BSM). The extraordinarily small magnitude of neutrino masses questions our understanding of mass generation for elementary particles, and may be an indication of a different origin for neutrino masses, beyond the Dirac mass model.

Currently, the leading theoretical candidate to explain neutrino masses is the so-called *Seesaw mechanism* [1–5]: the smallness of the observed neutrino masses is due to the existence of a new heavy stable particle called the *heavy neutral lepton* (HNL). Such hypothetical particles are singlets under each gauge group of the SM, therefore they cannot interact with SM particles through electroweak or strong interactions. Hence, they are said to be *sterile*. However, they can be produced through other mechanisms, such as mixing with SM neutrinos. According to the Seesaw mechanism, HNLs would give rise to gauge-invariant mass terms for the SM neutrinos.

The goal of this Thesis is to present an overview of the experimental and statistical methodologies employed to search for long-lived HNLs with the CMS experiment, as well as the results of the data analysis process. Previous analyses [6–8] in HEP typically focused on searching for HNLs with relatively high masses ($\mathcal{O}(100\text{ GeV})$ - $\mathcal{O}(1\text{ TeV})$), leaving prompt signatures in the detectors. Since no evidence was found for HNLs in that mass range, the interest of researchers slowly shifted towards long-lived particles leaving signatures that are displaced from the primary proton-proton interaction point. The search presented in this Thesis investigates HNL models with relatively lighter masses of 1-24 GeV, therefore having longer lifetimes. Due to their long lifetimes, the HNL models studied in this Thesis result in displaced signals in the CMS detector. The experimental signatures investigated in this search are the presence of a prompt, well-isolated lepton, which the events are triggered on, as well as a displaced lepton associated with a displaced jet, resulting from the decay of the long-lived HNL. Long-lived signatures come with their own challenges, as general-purpose experiments, such as *CMS experiment* at CERN, were originally designed to capture prompt physics kinematics. The main challenges of capturing displaced signatures include the difficulty of designing triggers, reconstruction and identification algorithms, as well as controlling SM background processes leaving similar signatures in the detector.

Machine learning (ML) has become increasingly popular in HEP in the past few decades, this research field being one of the first to apply such techniques at various stages in data analyses searching for new particles or performing precision measure-

ments. ML algorithms proved rather promising in searches for long-lived particles leaving displaced signatures in detectors, typical for BSM physics. Identification of particle jets by their flavour origin (*tagging*) is of particular importance to many searches for BSM physics. A novel *deep neural network* (DNN)-based jet tagger was successfully developed by the CMS Collaboration with the aim of conducting searches for long-lived BSM particles [9]. Consequently, the new displaced jet tagger was applied in the search for long-lived HNL particles in proton-proton (p-p) collision data collected with the CMS experiment between 2016-2018 described in this Thesis, consisting of an integrated luminosity of 137.19 fb^{-1} . Furthermore, a complementary *boosted decision tree* (BDT) event classifier was added to the analysis with the aim of separating HNL signal events from background, which was designed to be complementary to the displaced jet tagger algorithm.

The Thesis follows the structure listed below. Chapter 2 gives a thorough introduction to the theoretical foundations of the SM, as well as its shortcomings and proposed extensions, including the justification for the Seesaw mechanism and the potential existence of HNLs.

Chapter 3 presents the experimental setup of the LHC and the CMS experiment, summarising their technical working principles and structure of subdetectors. In addition, a brief introduction is provided to the trigger system employed during data collection, as well as its upgrade for the *High-Luminosity LHC* (HL-LHC). Chapter 4 provides an overview of the physics object and event reconstruction algorithms employed to process raw signals collected by the CMS detector into complex p-p collision event structures.

Chapter 5 describes the main statistical analysis techniques and methodologies employed throughout the search for long-lived HNLs, including Monte Carlo (MC) simulations of HNL signal and SM background processes; introduction to the concepts of the two ML algorithms used in the analysis process: BDTs and deep neural networks and statistical methods employed for uncertainty calculations and inference on hypothesis testing via likelihood methods and *upper limit* (UL) setting.

Chapter 6 describes the efforts of the Phase-2 upgrade process of the CMS Level-1 trigger (L1T) system for the HL-LHC. The HL-LHC will pose a particular challenge to the trigger system as the number of simultaneous interactions in p-p collision events will increase by an order of magnitude. To mitigate the increased level of *pile-up* (PU) events, information from the CMS detector's tracking system will be included in the L1T. Therefore, an end-to-end neural network was proposed to reconstruct the *primary vertices* (PVs) in p-p collision events, as well as correctly associate particle tracks with them.

Chapter 7 introduces the search for long-lived HNLs decaying semi-leptonically into a displaced jet and lepton. Besides adopting the new DNN displaced jet tagger, a novel aspect of the search is that it considers simultaneous couplings of HNL particles to all three generations of SM leptons. Chapter 8 explains the design process of the event-level binary classifier aiming to distinguish HNL signal events from SM background, using a BDT algorithm. Chapter 9 presents the evaluation of the efficiencies and scale factors applied in displaced lepton identification and reconstruction, as well as the treatment of other systematic uncertainties considered in the analysis. Finally, Chapter 10 presents the data-driven background estimation approach employed in predicting the yield of HNL signal and SM background events in the blinded signal region of the search. Comparing the expected results to those observed in data, ULs are set for a range of HNL benchmark scenarios.

Chapter 2

Theoretical background

2.1 The Standard Model of Particle Physics

Particle physics is the discipline of physics that studies elementary particles, as well as their mutual interactions through forces, aiming to understand the fundamental laws of nature. These fundamental particles constitute the building blocks of the Universe, currently known to be indivisible, with no internal structure. The mathematical formalism describing the laws governing these elementary particles and the associated fundamental forces is called the Standard Model of particle physics. A brief overview of this theory is presented in the following sections, based on Refs. [10–13]. Natural units are used throughout this chapter and electric charges are given in units of the electron’s charge.

2.1.1 Overview

The SM is the currently best-known and extensively tested theory that unifies three of the fundamental forces that govern elementary particle interactions: the *electromagnetic*, the *strong* and the *weak forces*, each being represented in fundamental interactions through an exchange of their corresponding mediator particle, called a *gauge boson*. These force carrier particles are the excitations of quantum fields acting on fundamental particles at interaction points. The mechanism of each fundamental interaction is described through the symmetry group of a *quantum field theory* (QFT): electromagnetic (EM) interactions are described by the theory of *quantum electrodynamics* (QED); strong interactions are described by *quantum chromodynamics* (QCD) and weak interactions are described through *quantum flavourdynamics*. The unification of the electromagnetic force with the weak force results in the *electroweak*

theory (EW) described by the unified Glashow-Salam-Weinberg (GSW) model [14–16]. Spontaneous electroweak symmetry breaking (EWSB) gives rise to the masses of fundamental particles through the Higgs mechanism [17, 18]. The mediator particle of the Higgs field is the Higgs boson, a prediction of the SM that was the final missing piece to validate the SM theory. Its discovery by the ATLAS and CMS Collaborations in 2012, resulted in the success of endorsing the SM as the fundamental theory of elementary particles and forces [19, 20].

These attractive forces, with different scales of strength, mainly govern at the quantum level of nature, while being negligible at macro-scales. The fourth fundamental force of Nature, gravity, however, has a negligibly weak effect at subatomic levels, whereas it becomes the main attractive force at large scales within the Universe, being responsible for its overall structure. To date, no satisfactory quantum theory of gravity has been established. Consequently, the gravitational force has not yet been successfully integrated into the SM.

Fundamental particles consist of spin-1/2 fermions, integer-spin force mediator gauge bosons and the spin-0 Higgs boson. Fermions are further grouped into quarks and leptons, both categories consisting of three generations of particles, where the quantum properties of the second and third generations are the exact copies of their first-generation cousins, except their masses being increasingly heavier. Each generation of fermions consists of two leptons or two quarks, respectively. The building blocks of ordinary matter consist of only first-generation particles, whereas their unstable, heavier versions decay through the mechanisms described by the fundamental interactions of particle physics. Eventually, all second- and third-generation particles decay to first-generation particles, which eventually group to form the constituents of more complex structures of matter, again, through the previously described fundamental forces.

Quarks may never be found as free particles in Nature, due to the strong force that governs them [21]. In fact, they are always found in bound states called *hadrons*, where quarks either group in pairs, called *mesons*, or groups of three, called *baryons*, according to the *quark model* of particle physics. However, at high energies, quarks may form even more exotic states, such as *tetraquarks*, *pentaquarks* and beyond, recently discovered by the LHCb collaboration [22, 23]. Top quarks, however, due to their large mass, do not hadronise before they decay, being the only "free quarks". Leptons, on the other hand, are always free particles, according to the SM. Quarks carry colour charge, a manifestation of their interaction with the strong force, whereas leptons do not have colour charge.

The dynamics of fundamental fermions are described by the *Dirac equation*, which

predicts the existence of *antiparticles*. In fact, each elementary particle has a corresponding antiparticle of the same mass but opposite-sign electric and parity charges. Therefore, when a fundamental particle interacts with its own antiparticle, they annihilate into energy in the form of a photon.

Table 2.1 summarises the properties of the fundamental fermions, including their electromagnetic charges and masses. Masses of charged leptons and of the top quark are obtained through high-precision measurements, whereas masses of other quarks are estimated via the $\overline{\text{MS}}$ -scheme [24]. Neutrinos are known to have non-zero masses, however, their exact values are yet to be determined. Currently, there are only upper limits available, which are observed to be at least nine orders of magnitude lower than other fermion masses. Apart from neutrinos, fermions of a particular generation have rather similar masses, which is thought not to be a coincidental aspect of the SM theory, but rather a low-energy manifestation of physics beyond the SM, as discussed in the next section (Sec. 2.2).

Table 2.1: Summary of the twelve fundamental fermions grouped by leptons and quarks, as well as by generation [24].

Generation	Leptons				Quarks			
	Particle (sym.)	Q		Mass	Particle (sym.)	Q		Mass
First	electron	(e^-)	-1	$\sim 511 \text{ keV}$	up	(u)	$+\frac{2}{3}$	$2.16^{+0.49}_{-0.26} \text{ MeV}$
	neutrino	(ν_e)	0	$< 1.1 \text{ eV}$	down	(d)	$-\frac{1}{3}$	$4.67^{+0.48}_{-0.17} \text{ MeV}$
Second	muon	(μ^-)	-1	$\sim 106 \text{ MeV}$	charm	(c)	$+\frac{2}{3}$	$1.27 \pm 0.02 \text{ GeV}$
	neutrino	(ν_μ)	0	$\lesssim 1 \text{ eV}$	strange	(s)	$-\frac{1}{3}$	$93^{+11}_{-5} \text{ MeV}$
Third	tau	(τ^-)	-1	$\sim 1.78 \text{ GeV}$	top	(t)	$+\frac{2}{3}$	$172.76 \pm 0.30 \text{ GeV}$
	neutrino	(ν_τ)	0	$\lesssim 1 \text{ eV}$	bottom	(b)	$-\frac{1}{3}$	$4.18^{+0.03}_{-0.02} \text{ GeV}$

Table 2.2 presents the properties of fundamental spin-1 vector bosons and their associated fundamental forces. Photons and gluons are massless force carrier gauge bosons while the electroweak bosons gain their mass via the Higgs mechanism by interacting with the scalar Higgs field. Consequently, an additional spin-0 boson emerges of mass $m_H = 125.25 \pm 0.17 \text{ GeV}$, called the Higgs boson. The masses of the Higgs and electroweak bosons are obtained from direct measurements [24]. In addition, table 2.3 shows which fermions experience the fundamental forces listed above. Since all charged fermions are massive particles, they all interact with the Higgs field via Yukawa interactions. Neutrinos, on the other hand, have very small, non-zero masses, currently not described by the SM. In fact, the SM predicts neutrinos to be massless particles that therefore do not couple to the Higgs field. More details

on potential neutrino mass generation mechanisms will be discussed in the next section (Sec. 2.2).

Table 2.2: Summary of SM fundamental forces with their associated mediator spin-1 gauge bosons, indicating the gauge symmetries that govern them. Since the Higgs boson is a spin-0 scalar boson, it is not a force carrier particle.

Force	Gauge group	Particle	Mass	Charge
Electromagnetic	U(1)	γ	0	0
Weak	SU(2)	$W^\pm$	$80.4 \pm 0.012 \text{ GeV}$	± 1
		Z	$91.1876 \pm 0.0021 \text{ GeV}$	0
Strong	SU(3)	g	0	0

Table 2.3: Table presenting the forces experienced by fermions.

		Strong	Electromagnetic	Weak
Quarks	up-type	$\checkmark$	$\checkmark$	$\checkmark$
	down-type			
Leptons	charged		$\checkmark$	$\checkmark$
	neutrino			

2.1.2 Gauge theories

Group theory is a discipline of mathematics that studies *symmetry groups*. A special branch of group theory is *Lie Theory* which deals with continuous symmetries. In the language of mathematics, a *group* is defined as a collection of transformations of a given set, G , and an associated binary operation, $*$, that satisfies a set of axioms. *Symmetry* is the invariance of an object under a set of transformations that eventually form a symmetry group [13].

Gauge theories correspond to symmetry groups of *fundamental fields* within the SM that obey invariance under *local gauge transformations*. Such transformations induce phase transformations to these particle fields with varying shift size at different points in spacetime. In contrast, *global gauge transformations* induce a constant phase shift to the fundamental fields at all spacetime coordinates, therefore preserving invariance under such transformations by definition. The mathematical formalism that describes the properties and interactions of fundamental particles within the SM obeys the Lie algebra of the symmetry groups they interact with. The Lie algebra is a vector space equipped with a Lie bracket operation which is a bilinear operation that takes two elements (vectors) from the vector space and returns another vector, satisfying properties, such as bilinearity, antisymmetry and the Jacobi identity. Consequently, the *generators* (or basis vectors) of the *Lie group algebra* generate

finite transformations and form the basis of the corresponding gauge fields, i.e. the mediator gauge bosons. If the generators of the Lie group commute, then such a group is said to be an *Abelian group*. While the SM is a *non-Abelian group*, its subgroup describing electromagnetic interactions is Abelian. Overall, the SM is a *Poincaré symmetry group* that is invariant under the spacetime transformations of special relativity, such as translations and rotations of spacetime coordinates, as well as transformations of gauge fields. The SM is the product of unitary and special unitary groups of gauge fields discussed in detail later in this section. A unitary group is defined as:

$$UU^\dagger = U^\dagger U = \mathbb{I} \quad (2.1)$$

where $U^\dagger$ is the Hermitian conjugate of unitary matrix U and $\mathbb{I}$ is the identity matrix. Special unitary groups $SU(N)$, besides obeying equation 2.1, also satisfy $\det(U) = 1$.

The Lagrangian formalism of the SM enables the translation of the abstract mathematical concepts of group theory into a physical context and aids working with symmetries in a convenient manner. The Lagrangian density, $\mathcal{L}(\phi, \partial_\mu \phi)$, describes the dynamics and interactions of fundamental particles in terms of quantum fields $\phi(x)$ and their $\partial_\mu \phi(x)$ spatial derivatives in spacetime coordinates ($\mu = \overline{0, 3} \equiv \{t, x, y, z\}$), adopting the 4-momentum notation: $x = (t, \vec{x})$. In fact, the Lagrangian, L , yields the equations of motions of such quantum systems. A Lagrangian may be decomposed into a kinetic energy term, T , that contains partial derivatives of fields and a potential energy term, V , that contains the mass, m , of particles that are excitations of the fields:

$$L = T - V = \int \mathcal{L} d^3\vec{x}. \quad (2.2)$$

In fact, the Lagrangian is the volume integral of the Lagrangian density. However, in the rest of the text, the latter will be referred to as the Lagrangian. The Lagrangian may be decomposed into a free field (i.e. freely propagating particles) and an interaction part involving multiple particles: $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{\text{int}}$. The action S of a system is defined as:

$$S = \int L dt = \iint \mathcal{L} d^3\vec{x} dt = \int \mathcal{L} d^4x. \quad (2.3)$$

Quantum fields follow the *principle of least action* ($\delta S = 0$), giving rise to the *Euler-Lagrange equations*:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \quad (2.4)$$

Another fundamental aspect of the SM is its consistency with Nöther's theorem [25], implying the existence of a conservation law associated with each continuous symmetry of a field theory under a gauge transformation. In other words, if the Lagrangian density stays invariant under a gauge transformation, then there is a conserved current from which one can construct conserved charges or quantum numbers. However, one can find discrete symmetries in Nature such as charge- (C), parity- (P) and time-conjugation (T). While their joint operation (CPT-conjugation) is a conserved symmetry of Nature, individual symmetries have been observed to violate conservation laws, for instance, the weak interaction has C-, P- and CP-violating effects. Charge conjugation for a left-handed neutrino would yield a left-handed antineutrino, which does not couple to the SM weak bosons. Therefore, the weak interaction maximally violates C-symmetry. Furthermore, based on the Sakharov conditions, to establish the matter-antimatter asymmetry observed in the Universe, C- and CP-symmetries must have been violated in the early Universe.

The Lagrangian of a free non-interacting scalar field may be written as:

$$\mathcal{L}_S = \frac{1}{2} [\eta^{\mu\nu} (\partial_\mu \phi) (\partial_\nu \phi) - m^2 \phi^2] \quad (2.5)$$

where $\eta^{\mu\nu} = \text{diag}(1, -1, -1, -1)$ is the Minkowski-space metric. An equation of motion for scalar fields is obtained by substituting $\mathcal{L}_S$ into the Euler-Lagrange equation 2.4, resulting in the *Klein-Gordon equation*. The latter is the relativistic generalisation of Einstein's energy-momentum relation ($E^2 = \vec{p}^2 + m^2$) written in a Lorentz-invariant formulation:

$$\eta^{\mu\nu} \partial_\mu \partial_\nu \phi + m^2 \phi = 0. \quad (2.6)$$

The Klein-Gordon equation permits negative-energy solutions (also arising from Einstein's equation) which is difficult to interpret in a physical context since it would predict a negative probability density. Moreover, since it originates from Einstein's relativistic equation for any particle, the Klein-Gordon equation should describe all particles observed in Nature. However, it has no solution for spin- $\frac{1}{2}$ fermion fields but only spin-0 bosons that can take the same quantum state, whereas fermions obey Pauli's exclusion principle. In the context of the SM, the Klein-Gordon equation is the free-field part of the Higgs boson.

In an attempt to solve the issue of negative-energy solutions of the Klein-Gordon equation, Paul Dirac sought to find an equation of motion that is linear in terms of the field and similar in form to the (non-relativistic) Schrödinger equation:

$$i\partial_t\phi(t, \vec{x}) = \left(-\frac{1}{2m}\nabla^2 + V(\vec{x}) \right) \phi(t, \vec{x}) \quad (2.7)$$

In fact, Dirac found a solution to the relativistic spin- $\frac{1}{2}$ fermion fields by substituting the Dirac Lagrangian, $\mathcal{L}_D$:

$$\mathcal{L}_D = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \quad (2.8)$$

into the Euler-Lagrange equation, obtaining the *Dirac equation*:

$$i\gamma^\mu(\partial_\mu\psi) - m\psi = 0 \quad (2.9)$$

where $\psi(x)$ is the spacetime-dependent fermion field expressed as a 4-component Dirac spinor, consisting of the spin-up and spin-down eigenstates of the positive energy and negative energy solutions to the equation. The positive energy solutions are interpreted as the particles while negative energy states represent antiparticles. The quantity γ^μ represents the Dirac gamma matrices that obey Clifford algebra with the anti-commutation relation $\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2\eta^{\mu\nu}$. The adjoint Dirac field is defined as $\bar{\psi} = \psi^\dagger\gamma^0$.

Quantum electrodynamics

Quantum electrodynamics (QED) is the relativistic formulation of electromagnetism and its Lagrangian, $\mathcal{L}_{\text{QED}}$, is required to stay invariant under local U(1) symmetry transformations: $\mathcal{L}_{\text{QED}} \xrightarrow{U(1)} \mathcal{L}'_{\text{QED}} = \mathcal{L}_{\text{QED}}$ [11]. Starting from the Dirac Lagrangian (Eq. 2.8), a free fermion field transforms under local U(1) as:

$$\psi \xrightarrow{U(1)} \psi' = e^{iq\alpha(x)}\psi \quad (2.10)$$

where q is the electric charge of the free fermion and $\alpha(x)$ is a spacetime dependent phase. Consequently, the Dirac Lagrangian transforms as:

$$\mathcal{L}_D \xrightarrow{U(1)} \mathcal{L}'_D = \mathcal{L}_D - q\bar{\psi}\gamma^\mu(\partial_\mu\alpha(x))\psi \neq \mathcal{L}_D \quad (2.11)$$

introducing a non-vanishing term in $\alpha(x)$. In order to restore local invariance, an interaction term is added to the QED Lagrangian, representing the interactions between charged fermions and photons:

$$\mathcal{L}_{\text{int}} = -q\bar{\psi}\gamma^\mu\psi A_\mu \quad (2.12)$$

where A_μ is the gauge boson (photon) field; in other words, the electromagnetic four-potential. This vector field transforms as:

$$A_\mu \longrightarrow A'_\mu = A_\mu - \partial_\mu\alpha(x) \quad (2.13)$$

restoring the gauge invariance of $\mathcal{L}_{\text{QED}}$ by cancelling the additional non-vanishing term in the Dirac Lagrangian of equation 2.11. To complete the QED Lagrangian, another gauge-invariant kinetic term is added, representing the freely propagating photon field:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (2.14)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the *field strength tensor* of the electromagnetic field. Absorbing the vector field A_μ into the *covariant derivative*, $D_\mu = \partial_\mu + iqA_\mu$, the complete, gauge-invariant QED Lagrangian can be expressed as:

$$\mathcal{L}_{\text{QED}} = \mathcal{L}_D + \mathcal{L}_{\text{int}} + \mathcal{L}_{\text{gauge}} = i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (2.15)$$

Applying the interaction and gauge terms of the QED Lagrangian to the Euler-Lagrange equations, one recovers Maxwell's equations. Since adding a gauge boson mass term ($\sim m_A^2 A^\mu A_\mu$) to $\mathcal{L}_{\text{QED}}$ would violate its local gauge invariance, the photon is predicted to be a massless vector boson. The importance of local gauge invariance lies in the fact that only such QFT models are renormalisable, hence predictive beyond a typically low-energy regime [26]. Consequently, the range of the photon field is practically infinite. Finally, based on Nöther's theorem, the conservation law associated with the QED Lagrangian obeying U(1) symmetry is the conservation of electric charge.

Quantum chromodynamics

While QED transforms under $U(1)$ with a single generator, i.e. a single force mediator boson (the photon), the strong and weak forces obey local gauge invariance via special unitary group transformations, $SU(N)$. Such groups have $N^2 - 1$ generators, giving rise to $N^2 - 1$ gauge bosons. Therefore, the gauge boson field may be rewritten in terms of the group generators t^a : $A_\mu = A_\mu^a t^a$, as well as the covariant derivative of $SU(N)$ becomes:

$$D_\mu = \partial_\mu + igA_\mu^a t^a \quad (2.16)$$

Consequently, the field strength tensor for a $SU(N)$ group can be also expressed in terms of the group generators:

$$\begin{aligned} F_{\mu\nu} &= -\frac{1}{g}[D_\mu, D_\nu] = \partial_\mu A_\nu - \partial_\nu A_\mu - g[A_\mu, A_\nu] = \\ &= (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a - gf^{abc}A_\mu^b A_\nu^c)t^a = F_{\mu\nu}^a t^a \end{aligned} \quad (2.17)$$

where f^{abc} are the antisymmetric structure constants of the Lie group defined with the commutation relation: $[t^a, t^b] = if^{abc}t^c$. If $f^{abc} = 0 \forall a, b, c$, then the generators are commutative and the symmetry group they serve basis to is an *Abelian group*, like the $U(1)$ Lie group of the electromagnetic force. However, if the generators of the Lie group do not commute ($f^{abc} \neq 0 \forall a, b, c$), the symmetry group is a *non-Abelian group* and implies self-interactions of the gauge fields. $SU(N)$ symmetry groups are non-Abelian, therefore their vector bosons self-interact.

Quantum Chromodynamics (QCD) is the fundamental theory of strong interactions between quarks and the mediator gauge bosons of the strong force, the gluons, described by the $SU(3)_C$ symmetry group [12]. Consequently, QCD is a non-Abelian gauge theory implying self-interactions among its gauge bosons. Its conserved quantity based on Nöther's theorem is the *colour charge* C . Gluons possess double colour charge and quarks have a single colour charge, while all other fundamental particles are colour neutral, therefore only quark fields interact with the gluon fields. *Colour confinement* is the key phenomenon observed in QCD that prevents the individual quarks and gluons from being directly observed as isolated particles due to their strong interactions. When quarks or gluons try to separate from one another, the energy between them increases, and new quark-antiquark pairs are created from the vacuum, forming new hadrons. In fact, only colour singlets are observed in Nature, hence quarks are always bound into hadrons either as a $q\bar{q}$ pair called *meson*

or as a $qqq/\bar{q}\bar{q}\bar{q}$ triplet called *(anti-)baryon* [21]. More exotic states of hadrons, such as tetraquarks and pentaquarks, have been also observed in high energy collisions at the LHC [22, 23]. As a matter of fact, baryon number, B , is also a conserved quantity in particle interactions: $B = N(\text{baryons}) - N(\text{antibaryons})$.

The Dirac spinor of a quark with flavour $f = u, d, s, c, b, t$ has three components: $\psi_f = (\psi_{f,\text{red}}, \psi_{f,\text{blue}}, \psi_{f,\text{green}})$ which transforms under local SU(3) as:

$$\psi_f \xrightarrow{SU(3)} \psi'_f = e^{igs\vec{\alpha}(x)\frac{\vec{\lambda}}{2}} \psi_f \quad (2.18)$$

where g_s is the coupling strength of the strong interaction, $\vec{\alpha}(x)$ is a three-dimensional phase and $\frac{\vec{\lambda}}{2}$ are the $3^2 - 1 = 8$ generators of the SU(3) symmetry group, represented by the 3×3 non-commuting Gell-Mann matrices. As a result, the QCD Lagrangian, $\mathcal{L}_{\text{QCD}}$, has the following for:

$$\mathcal{L}_{\text{QCD}} = \sum_{f=u,d,s,c,b,t} \bar{\psi}_f (i\gamma^\mu D_\mu - m_f) \psi_f - \frac{1}{4} G_{\mu\nu}^a G^{a,\mu\nu} \quad (2.19)$$

where m_f is the mass of the quark with flavour f and $G_{\mu\nu}^a$ are the 8 gluon field strength tensors. To keep the QCD Lagrangian gauge invariant, no mass term $\sim m_g^2 G_{\mu\nu}^a G^{a,\mu\nu}$ can be added to it, which means gluons must be massless gauge bosons. Since rotations of left and right-handed fermion fields in the left and right isospaces are invariant, the QCD Lagrangian has chiral symmetry $U(2)_L \times U(2)_R$. The last (gauge) term of the Lagrangian describes the dynamics of gluon self-interactions in QCD, gluon fields possessing double colour charge. This has the consequence that the strong force between gluons becomes increasingly large at smaller distances, called *vacuum screening*. The strong coupling constant $\alpha_S = g_S^2/4\pi$ is said to be a *running coupling*, indicating its changing strength by the scale of energy transfer in a given QCD interaction. The strength of the coupling constant decreases as a function of the energy scale where the interaction takes place, eventually leading to *asymptotic freedom* at very high energies. At energies achieved in collider experiments, the strong force becomes sufficiently weak to treat quarks and gluons as quasi-free particles called *partons*. Using perturbation theory, high energy hadron-hadron collisions are analysed in terms of single parton-parton collisions.

Parton distribution functions (PDFs) quantify the probability of finding a specific parton with a given fraction, x , of the proton's total momentum (or energy) at a given energy scale, μ_F , of a hadron-hadron interaction. The dimensionless quantity x is also called the *Björken scaling variable*. The momentum contributions from all partons with flavour, f_i , result in the following normalisation equation:

$$\sum_f^{\text{partons}} \int_0^1 \text{PDF}(x, f_i, \mu_F^2) \cdot x dx = 1, \quad (2.20)$$

where the *factorisation scale* μ_F defines the boundary between long-distance, non-perturbative low-energy physics and the short-distance, high-energy hard-scattering interactions where hadron constituents can be distinguished into partons. Below the factorisation scale, soft gluon emissions are absorbed into the PDF as they are treated as part of the internal proton structure. Above the factorisation scale, parton emissions in hadron interactions are treated as part of the hard-scattering process.

Extrapolating from experimental data, such as scattering experiments involving protons and neutrons, PDFs may be evaluated for hadrons revealing their internal structures at high energy scales (> 1 GeV) and the fraction of the hadron's total momentum carried by individual partons. Consequently, it has been observed that besides the *valence quarks* (those that form pairs and triplets as building blocks of hadrons), virtual gluons and *sea quarks* also carry significant fractions of the total energy of hadrons at high energies [27]. Once gluon fields reach sufficiently high energies, exceeding the mass of a quark-antiquark pair, such particle pairs may be created from vacuum. However, due to *colour confinement*, they quickly form composite colour singlet particles via *hadronisation* processes [21]. Moreover, in high-energy particle collisions such hadronisation effects result in outgoing highly collimated cascades of hadrons called *jets*.

On the other hand, the strong coupling constant increases rapidly at low energies. At energy scales as low as ~ 300 MeV, α_S reaches unity and calculations become non-perturbative, leading to divergences in infinite series calculations of observables. As a consequence, it becomes impossible to directly calculate the internal structures of hadrons at such energies and lattice field theory is used in calculations, instead. Therefore, QCD is considered to be a short-range force, but rather strong at nuclear scales.

Electroweak interactions

Weak interactions are based on the non-Abelian SU(2) symmetry group, where the conserved charge is the third component of the weak isospin, T_3 . Based on the outcome of the Wu experiment [28], where maximal parity violation was observed in charged current (CC) interactions, the weak force is required to have a V-A (vector minus axial vector) structure, unlike the EM or strong force. This means that in a spatial mirroring of $x \rightarrow -x$, weak interactions behave differently. In fact, fermion

fields can be decomposed into their left (L) and right-handed (R) chirality eigenstates: $\psi = \psi_L + \psi_R$ by using chiral projection operators: $P_{L,R} = \frac{1}{2}(1 \mp \gamma^5)$, where $\gamma^5 = i \prod_{i=0}^3 \gamma^i$. The consequence of the V-A structure of the weak interaction is that only left-handed fermions (ψ_L) and right-handed anti-fermions ($\bar{\psi}_R$) couple to the charged weak bosons, $W^\pm$, forming weak isospin doublets: $(T, T_3) = (\frac{1}{2}, \pm\frac{1}{2})$, the group symmetry often being denoted as $SU(2)_L$. Right-handed fermions and left-handed anti-fermions form isospin singlets with the charged weak bosons: $(T, T_3) = (\frac{1}{2}, 0)$. A common notation used for the left-handed fermion doublets and right-handed singlets is:

$$\begin{aligned} \psi_L : \quad L_{iL} &= \begin{pmatrix} \nu_{iL} \\ \ell_{iL} \end{pmatrix} & Q_{iL} &= \begin{pmatrix} u_{iL} \\ d_{iL} \end{pmatrix} \\ \psi_R : \quad \ell_{iR}, \nu_{iR}, u_{iR}, d_{iR}. \end{aligned} \quad (2.21)$$

where ℓ and ν denote the $i = 1, 2, 3$ generations of charged leptons and neutrinos, respectively, while u and d denote the up- and down-type quarks. To date, no right-handed neutrinos or left-handed anti-neutrinos have been observed in nature. Moreover, since they are colourless fermions with neutral electric charge, they do not couple to the strong or EM forces. Therefore, such neutrinos seem not to participate in any SM interaction and are considered *sterile*. However, there is no fundamental reason for right-handed neutrinos and left-handed anti-neutrinos not to be included in the SM. Historically, neutrinos were thought to be massless particles and their contributions were expected to vanish in the Yukawa component of the SM Lagrangian (see Sec. 2.1.3). However, neutrino oscillations proved that neutrinos do have mass, even if it seems to be orders of magnitude smaller than the other fermions. Further discussion about the neutrino mass hierarchy problem and mass generation mechanism is left for Sec. 2.2.

The gauge transformation of the $SU(2)_L$ symmetry group takes the following form

$$\psi \xrightarrow{SU(2)_L} \psi' = e^{ig\vec{\alpha}(x)\frac{1}{2}\vec{\sigma}}\psi \quad (2.22)$$

where $\vec{\sigma}$ are the 3 Pauli spin matrices, which are the group generators of $SU(2)_L$ and g is the coupling strength of the weak interaction. Consequently, the weak interaction has 3 gauge fields, W_μ^a .

Electroweak (EW) unification is based the Glashow-Salam-Weinberg (GSW) model [14–16] which consists of an internal symmetry between QED and the weak interaction: $SU(2)_L \otimes U(1)_Y$, where Y is the weak hypercharge, the equivalent of electric charge in the EW theory. The $U(1)_Y$ gauge symmetry transforms as:

$$\psi \xrightarrow{U(1)_Y} \psi' = e^{ig'\frac{1}{2}Y\beta(x)}\psi \quad (2.23)$$

where g' is the coupling strength associated with the $U(1)_Y$ symmetry and $\frac{1}{2}Y$ is its single generator. The gauge field analogous to A_μ in the $U(1)_Y$ symmetry group is B_μ . Thus, the EW local gauge transformation obeys:

$$\psi \xrightarrow{SU(2)_L \otimes U(1)_Y} \psi' = e^{ig\vec{\alpha}(x)\frac{1}{2}\vec{\sigma}} e^{ig'\frac{1}{2}Y\beta(x)}\psi \quad (2.24)$$

The defining conservation relation of the EW unification is:

$$q = T_3 + \frac{1}{2}Y \quad (2.25)$$

giving a relation between the electric charge, isospin and weak hypercharge of a particle. The EW Lagrangian can be written as:

$$\mathcal{L}_{\text{EW}} = i\bar{\psi}\gamma^\mu D_\mu\psi - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^a W^{a,\mu\nu} \quad (2.26)$$

where $B^{\mu\nu}$ and $W^{a,\mu\nu}$ are the field tensors of the B_μ and W_μ^a gauge fields. It can therefore be seen that the last two terms of the Lagrangian are the gauge terms, indicating self-interaction for the W_μ^a gauge fields via the non-commuting relations between the Pauli spin matrices. The covariant derivative D_μ takes the following form:

$$D_\mu = \partial_\mu + ig'B_\mu\frac{1}{2}Y + igW_\mu^a\frac{1}{2}\sigma^a. \quad (2.27)$$

Given the possible chirality states of fermions involved in the EW interaction (Eq. 2.21), the fermionic interaction term of the EW Lagrangian may be decomposed into its chirality states:

$$\mathcal{L}_{\text{fermionic}} = i\bar{\psi}_L\gamma^\mu D_\mu^L\psi_L + i\bar{\psi}_R\gamma^\mu D_\mu^R\psi_R \quad (2.28)$$

where

$$\begin{aligned}
D_\mu^L &= \partial_\mu I_2 + ig' B_\mu \frac{1}{2} Y_L I_2 + ig W_\mu^a \frac{1}{2} \sigma^a \\
D_\mu^R &= \partial_\mu + ig' B_\mu \frac{1}{2} Y_R.
\end{aligned}
\tag{2.29}$$

The physical bosons associated with the EW theory are extracted from the relations above, where the charged-current (CC) mediators of the weak interaction are a linear combination of $W_\mu^{1,2}$:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2) \tag{2.30}$$

whereas the neutral-current (NC) propagator, the Z boson and the photon are linear combinations of W_μ^3 and B_μ :

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \tag{2.31}$$

where $\theta_W = \arctan \frac{g'}{g}$ is the weak mixing (Weinberg) angle, measured directly by experiment to be $\sin^2 \theta_W \approx 0.22$ [24]. It has been established by experiment [29], that Z bosons also violate parity.

Despite the EW Lagrangian being invariant under local gauge transformation, it does not describe the observed universe correctly as it does not contain mass terms for mediator gauge bosons $\propto m_{W/Z}^2 W^{a,\mu} W_\mu^a$, or mass terms for fermion fields $\propto m_\psi (\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L)$, with the latter having different transformation properties for the L and R fermion fields. While photons are indeed massless bosons, as are gluons, the $W^\pm$ and Z^0 bosons have been measured to be massive [30, 31]. In addition, many fermions have been observed to have non-zero masses [32] (including neutral leptons, the neutrinos which are assumed to be massless by the SM). Therefore, in order to make the SM a self-consistent theory, a new mechanism is needed that can explain, that explains the masses of observed elementary particles, which is discussed in the next subsection.

2.1.3 Electroweak symmetry breaking and fermion mass generation

In order to make the SM a self-consistent theory that can describe all fundamental interactions and particles observed in nature, including the mass generation mecha-

nism of elementary particles, the Higgs mechanism is introduced to spontaneously break the symmetry of the electroweak interaction within a local gauge symmetry [17, 18]. In fact, local gauge symmetry of the SM is still preserved, while the lowest energy (vacuum) state of the EW symmetry ($SU(2)_L \otimes U(1)_Y$) is broken while still obeying a different local gauge symmetry ($U(1)_{EM}$) explained via the electroweak symmetry breaking (EWSB) mechanism:

$$\underbrace{SU(2)_L \otimes U(1)_Y}_{\text{EW}} \xrightarrow{\text{Higgs}} U(1)_{EM}. \quad (2.32)$$

The SM Higgs mechanism introduces an $SU(2)$ doublet of complex scalar fields with 4 additional degrees of freedom, a weak hypercharge of $Y_H = 1$ and a weak isospin of $T_3 = -\frac{1}{2}$:

$$H = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (2.33)$$

where H is known as the Higgs field and one of two complex scalar fields is required to be charged (ϕ^+), while the other to be neutral (ϕ^0) in order to correctly describe the physical weak bosons in the Higgs mechanism. The Higgs field must remain invariant under the following Lagrangian:

$$\begin{aligned} \mathcal{L}_{\text{Higgs}} &= (D_\mu H)^\dagger (D^\mu H) - V(H) \\ &= (D_\mu H)^\dagger (D^\mu H) - \mu^2 H^\dagger H - \lambda (H^\dagger H)^2 \end{aligned} \quad (2.34)$$

where D_μ has the same form as Eq. 2.27 and $\mathcal{L}_{\text{Higgs}}$ obeys global $U(1)$ symmetry. For the $V(H)$ potential energy term to have finite minima, λ has to be positive. However, there are no limitations on the values taken by μ^2 . If $\mu^2 > 0$, the $V(H)$ potential is fully symmetric around $H = 0$, obeying local gauge symmetry. However, if $\mu^2 < 0$, the local $U(1)$ symmetry is broken and the $V(H)$ potential can take infinite minima values: $\sum_{i=1}^4 \phi_i^2 = v^2 = -\frac{\mu^2}{\lambda}$ with one ground state realised, where v denotes the non-zero Higgs field vacuum expectation value (VEV). A common way to denote this is as follows:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (2.35)$$

Considering small perturbations around the vacuum state of the Higgs field and

introducing the *unitary gauge* in order to eliminate unphysical massless Goldstone bosons (corresponding to each broken generator in the EWSB mechanism), the Higgs doublet field becomes:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (2.36)$$

Inserting this form into $\mathcal{L}_{\text{Higgs}}$, the kinematic term involving the EW covariant derivative yields the weak boson mass terms:

$$m_W = \frac{1}{2}gv = 80.4 \text{ GeV} \quad m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} = 91.2 \text{ GeV} \quad (2.37)$$

as well as confirming the masslessness of the photon: $m_A = 0$. The strong interaction of the $\text{SU}(3)_C$ is unaffected by the EWSB mechanism, which means the gluon fields, G_μ^a , do not couple with the newly introduced scalar fields and remain massless. Consequently, by introducing the scalar Higgs field into the framework of the SM to break the EW symmetry, a mechanism describing the mass generation of weak gauge bosons is obtained while showing consistency with the masslessness of the EM and strong forces' gauge bosons.

The masses of the weak gauge bosons need to be measured experimentally since the values of the Higgs VEV and the coupling strengths of the EW interaction (g, g') are free parameters of the SM. As the mediating gauge bosons of the weak force are massive particles, the associated coupling strength appears weaker than that of the EM and strong forces at low energies, as well as operating over short distances limited by the masses of the weak bosons. However, at higher interaction energies, the strength of the weak force can be comparable with the QED interaction.

The value of the Higgs VEV can be determined in terms of the Fermi coupling constant, G_F :

$$v = (\sqrt{2}G_F)^{-\frac{1}{2}} \approx 246 \text{ GeV}. \quad (2.38)$$

The quadratic terms in $h(x)$ of the $\mathcal{L}_{\text{Higgs}}$ predict the existence of a scalar boson of mass $m_H = \sqrt{2\lambda}v$, depending on the VEV and the free parameter λ , which is the coupling constant of the scalar Higgs field. The Higgs boson was discovered in 2012 by the ATLAS and CMS experiments at CERN [19, 20], validating the last missing piece of the SM. Its mass has been measured to be $m_H = 125.25 \pm 0.17 \text{ GeV}$ [32].

Yukawa interactions

The mechanism through which fermions acquire mass after EWSB is incorporated in the SM Lagrangian via the Yukawa interaction term. As discussed above, only particles that couple to the Higgs field gain mass, whereas massless particles, such as photons and gluons, do not. Originally, neutrinos were thought to be massless particles in the SM, therefore, neutrinos were omitted from the Yukawa interaction term of the SM Lagrangian in its original form.

The fermion mass term in the SM Lagrangian must remain invariant under local $SU(2)_L \otimes U(1)_Y$ transformation and has the form:

$$\begin{aligned}\mathcal{L}_{\text{Yukawa}} &= -g_f(\bar{L}HR + \bar{R}H^cL) \\ &= -\sum_{i,j}(G_\ell^{ij}\bar{L}_{iL}H\ell_{jR} + G_D^{ij}\bar{Q}_{iL}Hd_{jR} + G_U^{ij}\bar{Q}_{iL}H^cu_{jR})\end{aligned}\quad (2.39)$$

where g_f is the Yukawa coupling to fermions, f , and L, R are the $SU(2)$ left-handed doublets and right-handed singlets, respectively. H is the Higgs doublet field and H^c is its conjugate. The extended form of the Yukawa interaction contains the fermion fields as given by Eq. 2.21 and $G_{\ell,D,U}$ are the Yukawa coupling matrices between multiple fermion generations in the charged lepton, down- and up-type quarks. G_ℓ is a diagonal matrix, hence flavour violation cannot be observed in the charged lepton sector of the SM, G_D is also diagonal for similar reasons. G_U is the 3×3 Cabibbo-Kobayashi-Maskawa (CKM) quark-mixing matrix whose (partially correlated) elements contain 4 free SM parameters that need to be measured experimentally. An important feature of $\mathcal{L}_{\text{Yukawa}}$ is the lack of a neutrino term, as it would require the existence of right-handed neutrinos, and no right-handed neutrinos have been observed in nature thus far. Therefore, the default construction of the SM assumes neutrinos to be massless fermions. Since neutrinos have been observed to be massive particles with masses orders of magnitude lighter than those of other fermions, a different mechanism needs to be constructed to describe neutrino mass generation, as well as to explain the large discrepancy between their masses and those of other SM particles (for further details, see Sec. 2.2).

Substituting the unitary gauge H doublet (Eq. 2.35) into $\mathcal{L}_{\text{Yukawa}}$, introduces terms in v that represent the fermion masses, while those containing $h(x)$ represent the fermion fields contributing to the Higgs boson. Consequently, the masses of fermions must depend on the Yukawa coupling and the VEV:

$$m_f = \frac{vg_f}{\sqrt{2}}. \quad (2.40)$$

Therefore, lighter fermions have weaker Yukawa couplings. Since the Yukawa couplings are free parameters of the SM, fermion masses have to be measured experimentally.

2.2 Physics beyond the Standard Model

This Section discusses several open questions of the Standard Model, as well as phenomena observed in Nature that seem to deviate from the current theory. Several extensions to the SM are discussed, particularly those that are relevant to the search on long-lived heavy neutral leptons described later in this Thesis. These extensions have the potential to accommodate some of the theoretical and/or experimental puzzles of the SM.

2.2.1 Introduction

Even though the SM has passed many tests over the years with flying colours and all its predictions have eventually been confirmed, there are still a number of fundamental open issues that it either fails to give any information about or contradicts based on its initial form of construction. A number of theories have been constructed to extend the SM such that it could explain such phenomena that are known to be beyond the SM (BSM). These BSM theories may even shed light on the nature of the SM theory itself, as well as solve the theoretical puzzles it currently poses. In addition, BSM phenomena may hide New Physics (NP) that the high energy physics (HEP) community is currently seeking. While physicists have, up until now, aimed to prove the validity and robustness of the SM, their current efforts are focused on trying to break it in various ways: precision measurements; direct or indirect searches for new BSM particles; searches for decay patterns violating the symmetries of the SM etc.

As mentioned in the previous Section, the SM seems to be a rather ad-hoc theory constructed based on profound theoretical ideas in order to fit experimental observations. The fact that it has 26 free parameters put in by hand, i.e. measured experimentally, instead of resulting from higher theoretical principles, may hint at the SM being essentially an effective theory, a low-energy manifestation of a higher

order theory, possibly even a Grand Unified Theory (GUT). Even though the SM having been confirmed and rigorously tested up to the EW scale, it may still break down at some higher energy scale, below the Planck scale, $\Lambda_{Planck} \sim 10^{19}$ GeV, which is about 15 orders of magnitude higher than the currently accessible energies in a controlled laboratory environment at the Large Hadron Collider (LHC). Based on the running of the coupling constants of the three fundamental forces of Nature (setting gravity aside), there are strong reasons to believe that they may unify at some higher energy scale. In that case, the three symmetry groups of the SM may be unified in a similar manner to the GSW mechanism associated with EW unification and eventually described by a single, larger symmetry group under a joint theory. Furthermore, the mass of the Higgs boson is around the EW scale, far lighter than any naive prediction based on the assumption that the SM is a valid theory up to the Planckian scales. There is no reason to believe that the SM is the ultimate theory of Nature valid all the way up to Planckian energy scales that dominated the Universe during the Big Bang. Moreover, the failure to include gravitational interactions in the framework of the SM, as well as the weakness of the gravitational force at the quantum level are a clear sign that there must be different mechanisms at work at higher energy scales. A working theory is still yet to be developed that would fully describe quantum gravity.

In addition, there are further theoretical puzzles regarding the nature of the SM theory. There is no explanation for the three-generation structure observed both in the quark and leptonic sectors of the SM. Moreover, the masses of charged fermions belonging to any given generation seem to be closely related, which seems unlikely to be accidental. Nevertheless, there is a mass hierarchy structure between the three generations of charged fermions, spanning over 5 orders of magnitude. Furthermore, the strength of the coupling constants of the three gauge interactions are of similar orders of magnitude, as well. Another theoretical puzzle is the electric charge quantisation among fermions: from all possible real values, they take either integer or rather simple rational values. There may exist a hidden symmetry describing this pattern at currently unknown energy scales.

There are four "accidental" global symmetries unused in the construction of the SM. To date, baryon and lepton number conservation have been validated through multiple measurements, yet it is not strongly justified by the SM. The vast matter-antimatter asymmetry observed in the Universe is a sign that these global symmetries must be broken for a matter-dominated Universe to exist.

On top of all, neutrinos seem to violate the observations described above. Firstly, the SM was constructed under the assumption that neutrinos are massless fermions.

However, the observation of neutrino mixing by multiple experiments demonstrates that they do have mass. Nevertheless, their mass is extremely small, more than 6 orders of magnitude smaller than the lightest charged fermion, the electron. Through flavour mixing, they violate lepton number conservation, while their mass production mechanism must be different to the other fermions of the SM. In addition, the mass hierarchy in the neutrino sector is as yet unclear; they may be either Dirac or Majorana particles, or possibly even both simultaneously. Neutrinos are rather elusive particles and their measured properties are the first objective proof for the breaking of the SM, adding more open questions to it, rather than answering existing ones.

Another strong hint for NP comes from recent observations in the precision measurement sector via the anomalous magnetic dipole moment of the muon measured in the *Muon $g-2$ experiment* [33]. The current discrepancy between theory and the measured magnetic dipole moment of the muon is at 4.2σ deviations, which may be a sign of BSM physics contributions to the loop diagrams involved.

There is also further compelling evidence for BSM physics observed in the wider Universe, at macroscopic scales. Based on cosmological and astronomical observations, there must be a non-luminous, gravitationally interacting matter penetrating the Universe, referred to as *dark matter*. The nature of this form of matter is as yet unknown and is the subject of a wide range of searches and phenomenological proposals. In addition, there is evidence for the accelerated expansion of the Universe, which is attributed to so-called *dark energy*, not described by the laws of the SM. If dark energy was simply vacuum energy represented via a cosmological constant added to the SM Lagrangian, its predicted value would be 10^{120} larger than the one measured in HEP experiments. Moreover, there is no explanation within the SM regarding the cause of rapid *inflation* in the early Universe, which is observed by the homogeneity and isotropic nature of the current day Universe, as well as its flat horizon.

Neutrino masses

The observation of neutrino oscillations, the mixing between different neutrino flavours in flight, established that at least two of the SM neutrinos have mass. The evidence for non-zero neutrino masses confirmed by oscillation experiments is arguably the first compelling evidence for BSM physics measured in a laboratory environment. The upper limit on the neutrino masses has been established from a number of different experimental results, including results from cosmology [32, 34],

as well as direct measurements from the decay of Tritium [35] and the double beta decay of Ca^{48} . The extraordinarily small magnitude of neutrino masses questions our understanding of the mass generation of elementary particles and may indicate a different origin for neutrino masses beyond the Dirac mass model [36, 37].

Cosmological probes set an upper limit on the total mass of neutrinos to be at the sub-eV level [32]:

$$\sum_{i=1,2,3} m_{\nu_i} < 0.120 \text{ eV}, \quad (2.41)$$

while direct measurements of Tritium β -decays place an upper limit on the electron neutrino masses to be $m_e < 0.8 \text{ eV}$.

Assuming a 3-neutrino model, neutrino flavour states are linear superpositions of 3 neutrino mass eigenstates. The relationship between the former and the latter is governed by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, which is the equivalent of the CKM matrix for the quark flavour sector:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \quad (2.42)$$

This relationship allows the possibility for neutrinos to oscillate between different flavour states, while manifesting their mixed mass eigenstates in flight. Solar and atmospheric neutrino oscillation experiments, as well as reactor neutrino experiments, have the sensitivity to measure differences between two neutrino mass eigenstates, based on the initial flavour and energy spectrum of the emitted neutrinos. Furthermore, these experiments allow the measurement of various parameters within the PMNS matrix, yielding a relatively uniform distribution, as opposed to the one of the CKM matrix which has a roughly diagonal pattern. There is no explanation in the SM for these observations. In addition, both matrices allow the presence of a CP-violating phase, which has been successfully measured for the CKM matrix via various meson mixing and decays. Current neutrino oscillation experiments are still not yet fully sensitive to measure CP violation in the neutrino sector, which is one of the primary goals of future neutrino experiments. The current measurements of neutrino mass differences are as follows:

$$\Delta m_{21}^2 = m_2^2 - m_1^2 = (7.53 \pm 0.18) \times 10^{-5} \text{eV}^2, \quad (2.43)$$

$$|\Delta m_{32}|^2 = |m_3^2 - m_2^2| = (2.44 \pm 0.12) \times 10^{-3} \text{eV}^2. \quad (2.44)$$

Since only the absolute mass difference between mass eigenstates 2 and 3 is known, it is still not yet clear what the neutrino mass hierarchy is. If $\Delta m_{32}^2 > 0$, neutrinos follow a normal hierarchy ($m_{\nu_1} < m_{\nu_2} < m_{\nu_3}$), whereas if $\Delta m_{32}^2 < 0$, neutrinos follow an inverted hierarchy ($m_{\nu_3} < m_{\nu_1} < m_{\nu_2}$). Fig. 2.1 shows the smallness of neutrino masses in comparison to other SM fermions, as well as hints at a potential hidden symmetry between charged leptons arising from the similarity of the masses within each generation.

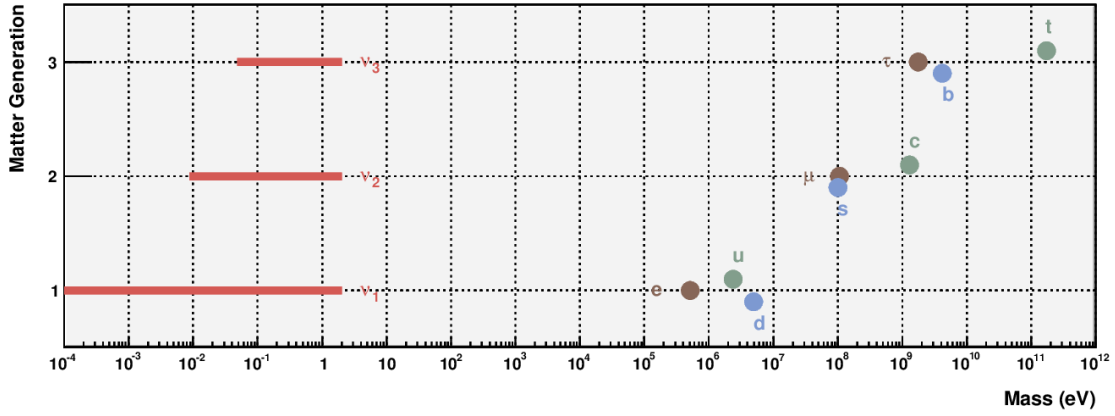


Figure 2.1: Fermion masses structured by their hierarchical order in 3 generations. The neutrinos are denoted as ν_i by their mass eigenstates and normal ordering is assumed. Since their exact masses are unknown, mass ranges are shown based on neutrino oscillation experiments. Figure taken from Ref. [38].

Two fixed-target, short-baseline neutrino oscillation experiments, LSND [39] and MiniBooNE [40], saw an excess of electron-neutrino and antineutrino appearance events at a short distance where oscillations from muon-neutrinos were not expected to occur. Such an excess was speculated to arise from a manifestation of a fourth-generation sterile neutrino. However, recent results from the subsequent MicroBooNE experiment [41] show no such excess and much of the sterile neutrino hypothesis' potential parameter space has been ruled out.

Nevertheless, the neutrino mass generation mechanism still remains a puzzle. The construction of the SM assumes neutrinos to be massless particles since the generation of Dirac mass terms in the SM Lagrangian assumes coupling between both left and right chiral states of fermions. Since the Goldhaber experiment [42] observed no neutrinos with right-hand helicity states, and no anti-neutrinos with left-hand

helicity states within the margin of error of the experiment, in the massless neutrino interpretation, this would have coincided with the non-existence of right-hand (left-hand) chirality states of neutrinos (antineutrinos). Therefore, no right-hand neutrino singlets would exist in the SM. As a consequence, no neutrino Dirac mass term would have been possible to be added into the SM Lagrangian, with the following form under spontaneous symmetry breaking:

$$\mathcal{L}_D = -m_D(\bar{\nu}_R\nu_L + \bar{\nu}_L\nu_R) \quad (2.45)$$

where R and L denote the right and left-hand chiral states of the neutrino. However, if neutrinos have mass, their helicity and chiral states may differ, which means right-handed neutrinos could still exist, even though it would be rather difficult to directly detect them. Even if right-handed neutrinos do exist and can be added to the SM Lagrangian in the usual way via a Dirac mass term, their extremely small mass would still not be justified, suggesting that a different mechanism may be needed to generate neutrino masses.

For example, the Majorana mass generation mechanism assumes that SM neutrinos consist of a superposition between both right- and left-handed chirality states, a similar approach to constructing the $SU(2)$ interactions of the weak theory:

$$\nu = \nu_R + \nu_L = C\bar{\nu}_L^T + \nu_L = \nu_L^C + \nu_L \quad (2.46)$$

where ν_L^C is the CP-conjugate field for the right-handed neutrino and corresponds to a left-handed antineutrino. Since neutrinos are neutral particles, it can be easily deduced that such Majorana neutrinos would be their own antiparticles:

$$\nu^C = (\nu_L^C + \nu_L)^C = \nu_L + \nu_L^C = \nu. \quad (2.47)$$

The Majorana mass term after symmetry breaking would have the following form:

$$\mathcal{L}_M = -\frac{1}{2}M(\bar{\nu}_L^C\nu_L + \bar{\nu}_R^C\nu_R), \quad (2.48)$$

constructed by mixing same-chirality fields. Since the left-chirality neutrino term is part of an $SU(2)_L$ doublet field, its Yukawa term would be inconsistent with the Higgs doublet model and would require the Higgs field to be extended into a triplet state. Therefore, Majorana mass terms cannot be added to the SM Lagrangian in a gauge

invariant manner and the lepton number would be violated in processes involving Majorana neutrinos. In this scenario, processes such as neutrinoless double- β decays would be observable in nature, a phenomenon currently being actively sought via dedicated HEP experiments [43].

In conclusion, the neutrino mass generation mechanism remains an open question since neither the Dirac, nor the Majorana models satisfy all the requirements for a gauge invariant theory consistent with the SM. Whether neutrinos are Dirac or Majorana particles is still a mystery, although there are BSM theories that would allow neutrinos to be both Dirac and Majorana particles simultaneously. Section 2.2.2 introduces one such theory that predicts the existence of heavier counterparts to the neutrinos observed in SM interactions. Such heavy sterile neutrinos are the subject of the search presented in this Thesis.

Dark Matter

Dark matter is perhaps the most compelling and probably first evidence for BSM physics, with its effects first being observed in the 1930s, even before the birth of the current SM theory. At the time, astrophysicists observed that, based on the dynamics and brightness of galaxy clusters, there must exist a non-luminous matter interacting gravitationally to account for the discrepancies between the amounts of visible matter and gravitationally interacting matter. It has been estimated that galaxy clusters consist of only about 2% visible matter of their total gravitationally interacting mass [44–46].

Another revolutionary discovery came in the 1970s with the measurement of galaxy rotational curves [47] showing large discrepancies between their theoretically proposed velocities using simple Newtonian gravity relations and those observed experimentally. the predicted velocity is given by:

$$v(r) = \sqrt{\frac{GM(r)}{r}} \quad (2.49)$$

where v is the rotational velocity of a galaxy at radius r from its centre, G is Newton's gravitational constant and M is the total mass of the galaxy within the radius r . Even accounting for additional mass arising from supermassive black holes lying in the centre of galaxies, the measured velocities could not be explained solely by the amount of visible matter (stars, dust, gas etc.). It was proposed that a halo of additional, invisible matter permeates throughout galaxies, making their outer edges rotate much faster than theoretically expected, based on the distribution of visible

matter. In fact, this halo of dark matter seems to be spherical, extending beyond the region of visible matter inside galaxies. Since this matter is not planar, unlike the typical shape of galaxies, it is assumed to be a collisionless form of matter, only weakly interacting with itself. These results appeared about 100 years after the first studies on velocity dispersions of stars orbiting the galactic centre of the Milky Way.

The existence of dark matter was further confirmed by gravitational lensing effects observed in galaxy clusters by the Hubble Space Telescope [48], where light emitted from far away galaxies bends due to extreme space-time curvature caused by the presence of non-luminous, massive bodies permeating the clusters, yet differing from black holes. Furthermore, the dynamics of galaxy cluster collisions show that while most galaxies pass each other without interaction, the hot gas permeating two clusters would indeed interact electromagnetically, causing friction and lagging. Nevertheless, similarly to the galaxies of the two colliding clusters, the dark matter permeating them does not interact, its motion remaining unaffected, as demonstrated by gravitational lensing effects.

Finally, the study of the Cosmic Microwave Background (CMB) led to astonishing results regarding the origin and the early evolution of the Universe. In the 1990s there were observations of anisotropies in the CMB spectrum by NASA's COBE satellite [49], which were subsequently studied in detail by NASA's WMAP [50] and the European Space Agency's (ESA) Planck satellites [51] in the first decade of the 21st century. The angular power spectrum observed by these experiments shows fluctuations in the CMB temperature over the sky, largely attributed to plasma fluctuations in the early Universe, before photons decoupled from matter and the Universe eventually cooled. The detailed structure of the angular power spectrum predicts the existence of gravitational matter clumps that did not interact with primordial photons and which behaved differently from baryonic matter. Moreover, the peaks in the CMB anisotropy can be used to predict the age (~ 13.8 billion years) and shape of the Universe (flat), along with its composition: 4.8% baryonic matter and radiation, 25.8% dark matter and 69% dark energy, a shocking result regarding our understanding of the Universe. It demonstrated that less than 5% of the constituents of the Universe are described by the SM. There was no knowledge of the nature and origin of the remaining 95% of the Universe's total mass-energy content; only its effect is known.

As a consequence, attempts to directly detect dark matter have been a priority for the HEP community in recent decades, although no evidence has been found for an elementary particle that would meet all the requirements for dark matter. The most popular dark matter candidate hypotheses assume it to be a non-baryonic,

massive particle that either interacts weakly or not at all with SM particles. Its nature can be categorised by its velocity, from cold (non-relativistic) to hot (highly relativistic) dark matter. Most popular theories support either cold or warm dark matter candidates. Sterile neutrinos are excellent candidates for both cold and warm dark matter.

Matter-antimatter asymmetry

One of the other fundamental mysteries of the Universe is its structure, which consists solely of matter. Based on standard cosmological models, equal amounts of matter and antimatter must have been produced in the early Universe after the Big Bang. Since the Universe was still constrained within a small space, matter and antimatter particles would continuously annihilate into photons which, in turn, would pair produce, leading to an equilibrium of matter-antimatter abundance. Therefore, there must have been a process in the early Universe that created a vast excess of matter over antimatter.

In the 1960s, Andrei Sakharov proposed a set of necessary conditions for baryogenesis to occur [52]:

- baryon number B violation;
- C- and CP-symmetry violation;
- interactions out of thermal equilibrium.

So far, no baryon number violating process has ever been observed in Nature. Baryon number violation can occur at non-perturbative levels of the SM at high energies via quantum tunnelling effects, called *sphalerons* [53]. However, the conditions such processes required to observe cannot be produced at present-day colliders. Sphalerons are saddle points in the EW potential, converting a triplet of baryons into a triplet of anti-leptons. Nevertheless, the difference between baryon and lepton number ($B - L$) is still an exact symmetry of the SM and, hence would be conserved. At the high temperatures of the early Universe prior to EWSB occurring, the abundance of sphalerons is expected to have wiped out any potential baryon excess.

C- and CP-symmetry violations have been observed in weak SM interactions, originally in the quark sector [54]. Both the CKM and PMNS matrices accommodate a CP-phase; however, the extent of CP violation observed in the quark sector is

insufficient to explain the asymmetry observed between matter and antimatter. CP-symmetry violation is yet to be confirmed in the neutrino sector, but it is one of the research goals of next-generation neutrino experiments [55]. Nevertheless, an additional source of CP violation must have existed in the early Universe.

Finally, CPT-symmetry, which is a fundamental symmetry of the SM, ensures that all processes compensate for one another, hence no baryon excess could eventually occur. The non-equilibrium condition is satisfied via the rapid inflation model of the early Universe, which proposed that different regions of the rapidly expanding Universe became separated at a higher rate than pair-annihilation between matter and antimatter could occur. Thus, the process generating baryon excess would have to occur at a slower pace than the expansion rate.

It can therefore be concluded that there must have been additional, yet unknown processes that occurred in the early Universe that, in turn, generated a net excess of baryonic matter that eventually led to a matter-dominated Universe. Nevertheless, it is still unclear what caused the primordial antimatter to disappear or BSM processes could have caused the antimatter to vanish.

2.2.2 The Seesaw mechanism

The so-called type-I *Seesaw mechanism* is a generic model used in the development of a gauge-invariant theory to explain the magnitude of neutrino masses [56–58]. The most generic, renormalisable Lagrangian mass term that could be integrated into the SM consists of a superposition of the Dirac and Majorana mass terms:

$$\mathcal{L}_{\text{D-M}} = -\frac{1}{2} \left[m_D \bar{v}_L v_R + m_D \bar{v}_R^C v_L^C + M \bar{v}_R^C v_R \right] + \text{h.c.} \quad (2.50)$$

This can be decomposed into matrix form as:

$$\mathcal{L}_{\text{D-M}} = -\frac{1}{2} \begin{pmatrix} \bar{v}_L & \bar{v}_R^C \end{pmatrix} \underbrace{\begin{pmatrix} 0 & m_D \\ m_D & M \end{pmatrix}}_{\mathbb{M}} \begin{pmatrix} v_L^C \\ v_R \end{pmatrix} + \text{h.c.} \quad (2.51)$$

While the Dirac mass term ensures the possibility of neutrino masses arising from the EWSB of the Higgs mechanism, the Majorana term adds a gauge-invariant component to the Lagrangian. The physical neutrino states are then the eigenvalues of the mass matrix $\mathbb{M}$ and can be obtained via diagonalising $\mathbb{M}$ in a similar manner

to the procedure used to determine the masses of the neutral gauge bosons, i.e. by solving the characteristic equation:

$$\det(\mathbb{M} - \lambda I) = 0. \quad (2.52)$$

This yields two solutions:

$$\lambda_{\pm} = m_{\pm} = \frac{M \pm \sqrt{M^2 - 4m_D^2}}{2} = \frac{M \pm M\sqrt{1 - 4(\frac{m_D}{M})^2}}{2}. \quad (2.53)$$

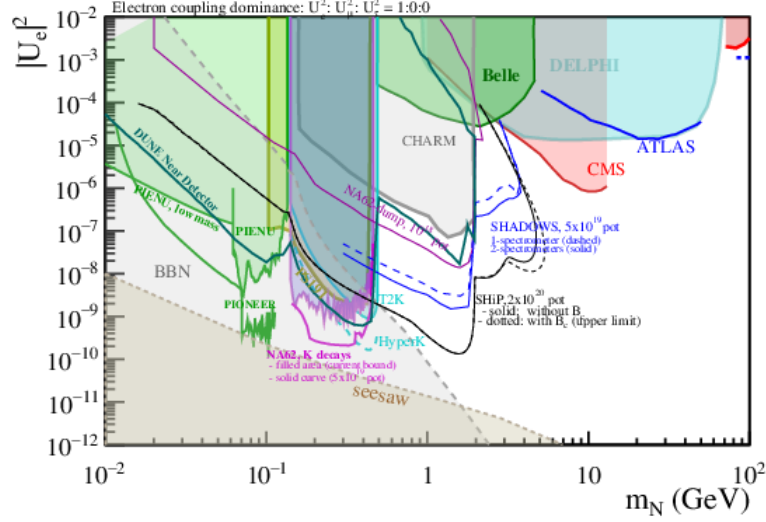
Based on the assumptions of the type-I Seesaw mechanism, the Majorana mass term M is far larger than the Dirac mass term m_D ($M \gg m_D$). As a result, Equation 2.53 can be further simplified using Taylor's expansion:

$$m_{\pm} \approx \frac{1}{2}M \pm \frac{1}{2}\left(M + 2\frac{m_D^2}{M}\right). \quad (2.54)$$

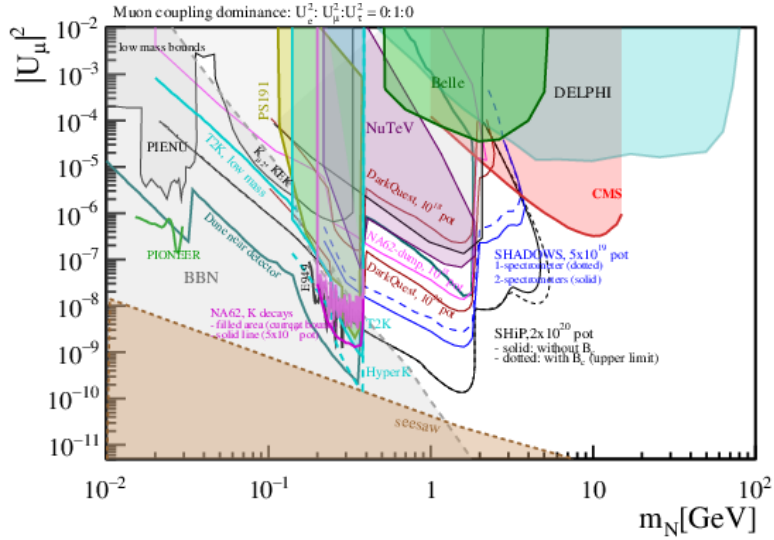
Consequently, two mass states are obtained: one for a light, "active" neutrino state (ν) that participates in SM interactions via the EW interaction and one for a heavy Majorana state neutrino, (N) that cannot participate in SM interactions, since it is mostly a right-handed massive neutrino and would only couple extremely weakly to the weak gauge bosons. They are therefore *sterile* fermions that can only mix with the active neutrinos of the SM. The mass eigenstates of the active and sterile neutrinos are:

$$m_- = m_{\nu} \approx \frac{m_D^2}{M} \quad \text{and} \quad m_+ = m_N \approx M. \quad (2.55)$$

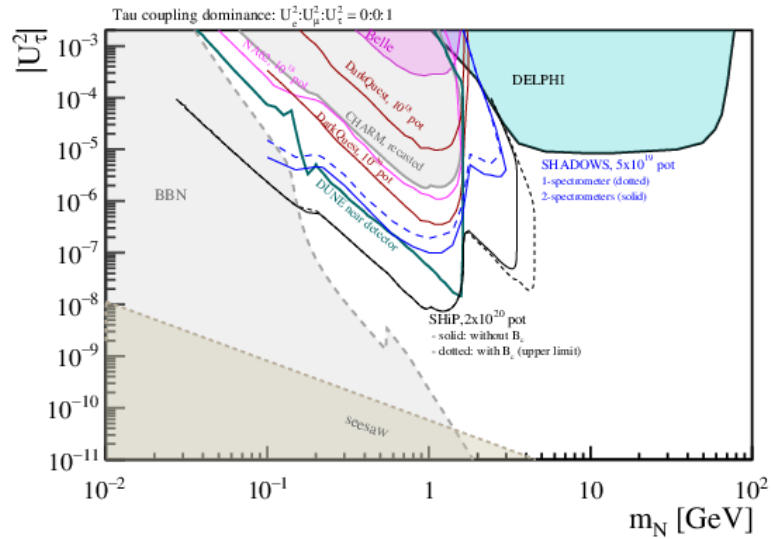
In the context of the Seesaw mechanism, the Dirac component of the neutrino masses must be around the same scale as the charged lepton masses, i.e. of the order of $\mathcal{O}(1\text{MeV}) - \mathcal{O}(1\text{GeV})$. In order to obtain the much smaller observed neutrino masses (around $\mathcal{O}(0.1\text{eV})$), the Majorana mass component must be large, that is, at around the GUT scale: $\sim \mathcal{O}(10^{15}\text{GeV})$. In fact, this may not be a coincidence, and the Seesaw mechanism could potentially further motivate the legitimacy of GUT theories. Nevertheless, testing this hypothesis at such energy scales remains an unachievable challenge at the present time. Assuming the Majorana mass to be generation independent, the neutrino masses would have the following relationships (assuming normal hierarchy): $m_{\nu_1} : m_{\nu_2} : m_{\nu_3} = m_e^2 : m_{\mu}^2 : m_{\tau}^2$.



(a)



(b)



(c)

Figure 2.2: Exclusion limits on the mixing between the electron-, muon- and tau-neutrino and a single heavy neutrino in the mass range of 10 MeV – 100 GeV, as well as the expected sensitivity from future experiments [59].

The *Neutrino Minimal Standard Model* (ν MSM) [60] has been proposed as a minimal extension to the SM to accommodate the non-zero neutrino masses. It predicts the existence of three heavy right-handed Majorana neutrinos, called Heavy Neutral Leptons (HNLs), denoted as N_1, N_2, N_3 , each one corresponding to the three fermion generations of the SM. These hypothetical particles are singlet fermions under each SM gauge group, hence they cannot interact with any SM particles or forces, but only mix with the active SM neutrinos. Eq. 2.55 predicts the HNL masses to be around the GUT scale, which makes this model rather difficult to test experimentally with current technology. There are, however, phenomenological models that predict HNLs with much lower masses, but that are still consistent with small Higgs couplings, such as the inverse and linear Seesaw models, even though such models are less natural for GUT-scale Majorana masses [59]. In fact, measurements of the top quark and Higgs boson masses revealed that the ν MSM could be a consistent effective theory up to very high energy scales, potentially right up to the Planck scale. The HNL candidates sought in this Thesis range between 1-24 GeV. Other searches, past and present, listed in Fig. 2.2 and Ref. [59], have sought HNL candidates with masses between ~ 10 MeV and the TeV scale.

The ν MSM is constructed in a similar manner to the EW theory introduced in Sec. 2.1.2, and is governed by the following symmetry groups: $SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L}$, where the generator of the $U(1)$ symmetry is $B - L$. The associated conservation law is therefore given by:

$$q = T_{3L} + T_{3R} + \frac{1}{2}(B - L). \quad (2.56)$$

If $SU(2)_L$ remains unbroken at high energy scales (assuming the validity of the SM up to GUT scales), but $SU(2)_R$ does not, the above relationship may be simplified to give:

$$\Delta T_{3R} = -\frac{1}{2}\Delta(B - L). \quad (2.57)$$

Eq. 2.57 introduces a potential source of baryon and lepton number violation while preserving the invariance of $B - L$ [61]. This could potentially explain the matter-antimatter asymmetry observed in the Universe via *leptogenesis*, which would, in turn, induce the *baryogenesis* required by the Sakharov conditions defined in Sec. 2.2.1. In fact, the Lagrangian in Eq. 2.51 would violate both baryon and lepton numbers, but their difference would remain constant, too, according to the generator of the $U(1)$ symmetry group. The mass matrix contains CP-violating phases and HNL decays

could include lepton number violating (LNV) processes and provide out-of-balance lepton excess. This, in turn, would result in an excess of baryons via anomalous fermion number non-conservation. The observed matter-antimatter asymmetry in the Universe places constraints on the extent of LNV effects and predicts that the two heavier HNLs, N_2 and N_3 , must have almost degenerate masses, as is the case with the active neutrinos (see Eq. 2.44): $|m_{N_2} - m_{N_3}| \ll m_{N_2}, m_{N_3}$. In addition, this phenomenon places limits on the mass range of the heavier sterile neutrinos of $0.15 \text{ GeV} \leq m_{N_{2,3}} \leq 100 \text{ GeV}$ [62]. Big Bang Nucleosynthesis (BBN) also places limits on the possible lifetimes of these heavier sterile neutrinos: $\tau_{2,3} < 0.2 \text{ s}$ [63].

The existence of HNLs may not only provide a reasonable framework for describing the origin and scale of neutrino masses (as well as being a source of baryogenesis), but they are also excellent potential candidates for warm dark matter particles, particularly their lightest version, N_1 . Based on theoretical models [64], the lightest sterile heavy neutrino would couple very weakly to SM neutrinos, as well as having a mass around $\mathcal{O}(1 - 10 \text{ keV})$ and a lifetime exceeding the age of the Universe. Its tiny Yukawa coupling would leave the lightest active neutrino almost massless. Being a stable particle and not interacting with other SM particles would make the lightest HNL particle an excellent dark matter candidate, even though it would be rather challenging to detect. Nevertheless, the two heavier HNLs, N_2 and N_3 , would be expected to have stronger couplings to active neutrinos, making their mixing with SM neutrinos a potentially viable approach to detect them. In fact, the couplings of N_2 and N_3 to active neutrinos are hypothesised to be roughly of similar strength and due to their approximately degenerate masses, the two heavier sterile neutrinos can be treated as a single HNL in experimental searches, denoted N . Consequently, the coupling strength or mixing angle between N and a SM neutrino of flavour α is denoted V_α , where $\alpha = e, \mu, \tau$.

HNLs can either couple to a single SM neutrino flavour or mix with multiple generations simultaneously. Previous HNL searches have mostly considered single-generation coupling scenarios, setting limits on a broad range of HNL couplings and masses ranging between 100 MeV and TeV scale, as presented in Fig. 2.2. There are also further constraints placed on the HNL mass-coupling phase space by the Seesaw mechanism and BBN, as already discussed earlier in this Section. The HNL lifetime is typically a backward oblique line in these diagrams, the lower the HNL mass, the more displaced it is for the same coupling. In fact, different strategies are needed to cover the broad spectrum of possible HNL coupling and mass scenarios, based on the accessible lifetimes to individual experiments, as well as their sensitivity, production and decay modes investigated. First of all, one needs to consider the possible production modes of HNL particles via mixing with active neutrinos in EW

interactions:

- W boson decays: $W^\pm \rightarrow \ell_{1,\alpha}^\pm \nu_\alpha \rightarrow \ell_{1,\alpha}^\pm N$;
- Z or Higgs boson decays: $Z^0, H \rightarrow \nu_\alpha \bar{\nu}_\alpha \rightarrow \nu_\alpha N$;
- Heavy- and light-flavour meson decays (if kinematically allowed, putting a limit on the probed HNL masses): $X^\pm \rightarrow \chi^0 \ell_{1,\alpha}^\pm \nu_\alpha \rightarrow \chi^0 \ell_{1,\alpha}^\pm N$,

where the subscript "1" in $\ell_{1,\alpha}^\pm$ indicates that it is a prompt lepton produced alongside the HNL, offering a feature that could be used for online event selection (trigger). $X^\pm$ indicates a charged meson decaying into a lighter neutral meson, χ^0 . HNL production via Z or Higgs boson decay offers a rather unattractive channel for searches as such decays would result in large missing energy events with no other features that could be exploited for online selection. Nevertheless, such processes are being investigated in current hadron collider data. HNL production via meson decay is kinematically limited to the mass of the parent meson - typically a few GeV - making the available HNL mass parameter space rather limited. In addition, the products from such meson decays have soft kinematics, making them rather challenging for online event selection in general-purpose detectors. Fig. 2.3 shows the Feynman diagrams associated with the HNL production channels discussed above.

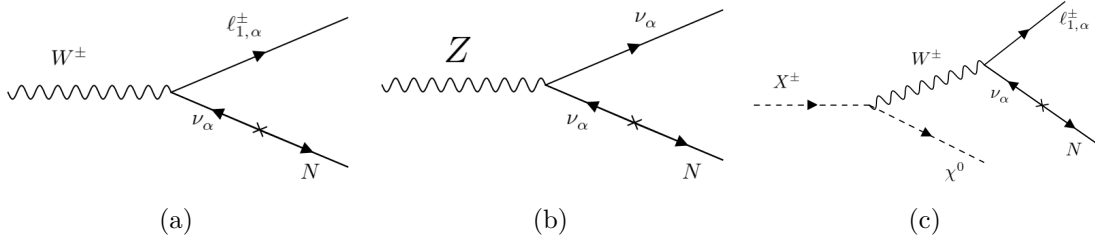


Figure 2.3: Feynman diagrams presenting various HNL production channels via mixing with active neutrinos of flavour $\alpha = e, \mu, \tau$. The left and middle diagrams show HNL production via EW gauge boson on-shell decays, while the diagram on the right shows HNL production via charged meson decay.

HNL decays take place exclusively via electroweak gauge bosons that then decay into SM particles. The analysis presented in this Thesis is a search for HNL particles decaying via off-shell EW bosons; however, if $m_N > m_{W,Z,H}$, the HNLs decay via on-shell gauge bosons. Based on the decay channel, such processes can be categorised into either leptonic or semi-leptonic final states:

- Leptonic final states:

- Via off-shell W^* boson: $N \rightarrow \nu_\beta \rightarrow \ell_{2,\beta}^\pm W^{*\mp} \rightarrow \ell_{2,\beta}^\pm \ell_{3,\gamma}^\mp \nu_\gamma$;
- Via off-shell Z^* boson: $N \rightarrow \nu_\beta \rightarrow \nu_\beta Z^* \rightarrow \nu_\beta \ell_{2,\gamma}^\pm \ell_{3,\gamma}^\mp$.
- Semi-leptonic final states:
 - Via off-shell W^* boson: $N \rightarrow \nu_\beta \rightarrow \ell_{2,\beta}^\pm W^{*\mp} \rightarrow \ell_{2,\beta}^\pm q \bar{q}'$;
 - Via off-shell Z^* boson: $N \rightarrow \nu_\beta \rightarrow \nu_\beta Z^* \rightarrow \nu_\beta q \bar{q}$.

The subscripts "2" and "3" indicate that the lepton is a daughter particle of the HNL and can be potentially displaced, depending on the lifetime of the HNL. The subscripts β and γ indicate the flavour of the neutrinos and leptons involved in the final state. The leptonic final states consist of three well-isolated leptons associated with some missing energy from the undetected neutrino. For leptonic decays via an off-shell Z^* boson, the two charged leptons originating from the HNL have the same flavour, while those coming from an off-shell W^* boson may have different flavours. Semi-leptonic final states are, however, fully visible and the HNL mass can be directly reconstructed from its decay products. In fact, semi-leptonic final states are particularly attractive for searches due to their higher decay cross-section, which is more than double that of the corresponding leptonic final states. Nevertheless, increased SM backgrounds may make it challenging to identify these final states. Lepton number violation can occur in collision events associated with HNLs if the HNL is a Majorana particle.

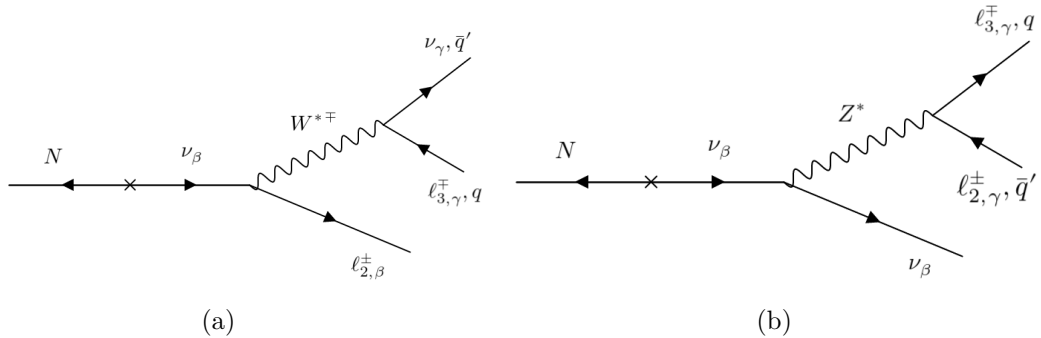


Figure 2.4: Feynman diagrams representing the leptonic and semi-leptonic HNL decay channels via mixing with active neutrinos of flavour $\beta = e, \mu, \tau$. The left diagram shows an HNL decaying via an off-shell W^* boson, while the diagram on the right shows an HNL decay via an off-shell Z^* boson. γ marks the flavour of the final state leptons resulting from the off-shell bosons.

Based on the extent of simultaneous HNL mixings with the three SM neutrino generations, its Dirac or Majorana nature could be revealed if, in case of discovery, the HNL couplings lie in the red (normal hierarchy) or blue (inverted hierarchy)

regions of the coupling matrix shown in Fig. 2.5. The lines in the matrix are normalised with respect to all three lepton flavours: $V_\alpha^2/(V_e^2 + V_\mu^2 + V_\tau^2)$. The elements of this coupling matrix have been determined by fitting current neutrino oscillation data, EW precision measurements and existing exclusion plots (Fig. 2.2). In addition, identification of the HNL coupling strengths would indirectly reveal the value of the leptonic CP-phase. This would be a significant breakthrough since the leptonic CP-phase embedded in the PMNS matrix has been experimentally challenging to measure and would allow the completion of the measurements of all the elements of the PMNS matrix.

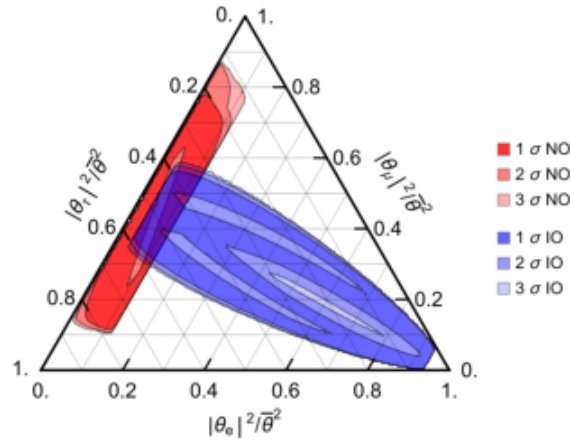


Figure 2.5: The range of normalised flavour mixings between HNLs and active neutrinos, consistent with the current neutrino oscillation data, assuming degenerate heavier sterile neutrino states. The different contours correspond to the allowed case of normal hierarchy (red) and inverted hierarchy (blue) within the shown uncertainties. Figure taken from Ref. [59].

2.3 Long-lived particles

This Section presents the general concept of a long-lived particle (LLP) and the motivation for long-lived HNL searches. While the SM contains long-lived particles with displaced signatures, such as displaced tracks or vertices, LLPs are also a broad and well-motivated sector of BSM physics. A brief overview of the opportunities and challenges in searching for LLPs with general-purpose detectors will be given.

2.3.1 Introduction

Long-lived particles consist of a family of elementary or composite particles that have long lifetimes resulting in a measurable displacement between their production

and decay points.

Historically, general-purpose detectors at the LHC have been designed to search for prompt, highly energetic particles, with the goal of discovering New Physics via such processes. However, in recent years there has been a shift of interest towards particles manifesting displaced signatures as potential candidates for BSM processes. Searches for such LLPs need different experimental approaches and analysis designs compared to prompt analyses, from dedicated triggers to sophisticated statistical interpretation of the results.

The proper lifetime of a particle is defined as:

$$\tau_0 = -\frac{t}{\ln \frac{N(t)}{N_0}} \quad (2.58)$$

where $N(t)$ is the average number of remaining particles after time t has passed and N_0 is the total number of particles at time $t = 0$. For LLPs, the lifetime can be interpreted as a distance, $c\tau_0$, assuming the LLP is travelling at roughly the speed of light. The total decay width of a particle is inversely proportional to its proper lifetime:

$$\Gamma = \frac{1}{\tau_0}. \quad (2.59)$$

The average distance travelled by a LLP particle of mass m and momentum p in the transverse plane of a detector from its production point is given by:

$$\langle L_{xy} \rangle = \sqrt{\frac{2}{3}} \beta \gamma c \tau_0 \equiv \sqrt{\frac{2}{3}} \frac{p}{m} c \tau_0 \quad (2.60)$$

where $\beta\gamma$ is its boost.

According to Fermi's golden rule, the partial decay width of a particle that transitions from state i to state f is:

$$\Gamma_{fi} = 2\pi |T_{fi}|^2 \rho(E_i), \quad (2.61)$$

where T_{fi} is the transition *matrix element* (ME) and $\rho(E_i)$ is the *density of states* at energy E_i . The transition ME can be expressed via the Lorentz-invariant ME, $\mathcal{M}_{fi}$, which, in turn, describes the dynamics of the interactions and is computed

from the appropriate Feynman rules. As a result, longer lifetimes correspond to small transition amplitudes from weaker couplings; smaller available phase space (either small mass split between particles in a decay chain or a multitude of final state particles) or heavy off-shell mediator particles involved in the decay process. Consequently, composite SM particles decaying via the weak force have longer lifetimes, compared to those undergoing decay via the strong force, as illustrated in Fig. 2.6. Hadrons decaying via the weak gauge bosons are classified as long-lived, whereas those decaying via the strong interaction have extremely small lifetimes.

In fact, particle lifetimes in the SM range over many orders of magnitude, from the shortest (Z boson $\sim 2 \times 10^{-25}$ s) to stable particles, such as the electron. In the SM protons are also considered stable, as they are the lightest composite baryons, yet a number of BSM models predict their decay. Searches so far set the proton lifetime limit beyond the age of the Universe ($\geq 10^{30}$ years) [32]. "Detector-prompt" particles have shorter lifetimes than the resolution of the detector, that is they decay close to their production point, which the detector cannot resolve. "Detector-stable" particles manifest themselves as apparently stable particles because their lifetimes are longer than the physical length of the detector, therefore they are unlikely to decay inside it. However, LLPs have lifetimes such that the distance they travel before decay is short enough to produce a displaced signature inside the detector in relation to their production point. The typical lifetime range for which a particle is considered an LLP is between $10 \mu\text{m}$ and 10 m .

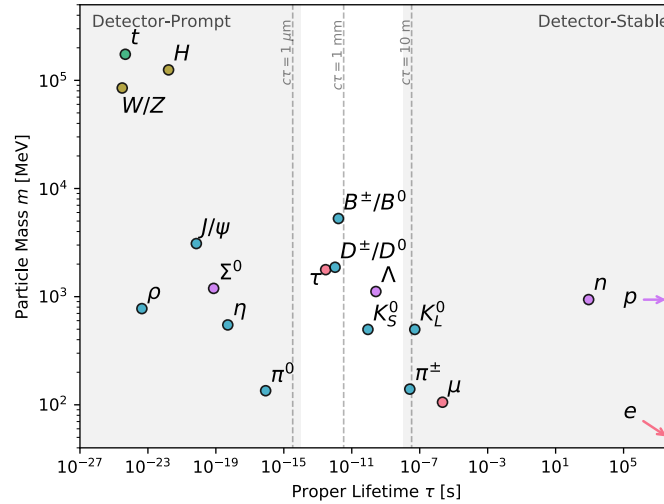


Figure 2.6: Particle masses and lifetimes for both elementary and composite SM particles. Prompt and stable particles are marked by the shaded regions, while lifetimes of $10 \mu\text{m}$ up to 10 m are associated with LLPs decaying inside a general-purpose detector. Figure from Ref. [65].

Typically, a particle can decay via multiple channels and the total decay width of such a particle is the sum of the associated partial decay widths:

$$\Gamma = \sum_j \Gamma_j. \quad (2.62)$$

Consequently, for a particular decay channel, the branching ratio is defined as:

$$\text{BR}(j) = \frac{\Gamma_j}{\Gamma}. \quad (2.63)$$

BSM theories often predict the existence of LLP particles over a wide mass and lifetime range. As shown in Eq. 2.61, new weak-scale particles can easily have long lifetimes due to their small couplings to SM particles, the suppressed phase space available for decays or the approximate symmetries that stabilise them. The experimental signatures of such hypothetical BSM particles in general-purpose detectors are varied, ranging from displaced tracks, vertices or jets, to disappearing or kinked tracks, showing rather different topologies from SM processes. Fig. 2.6 presents graphically the potential signals that may indicate sources of New Physics.

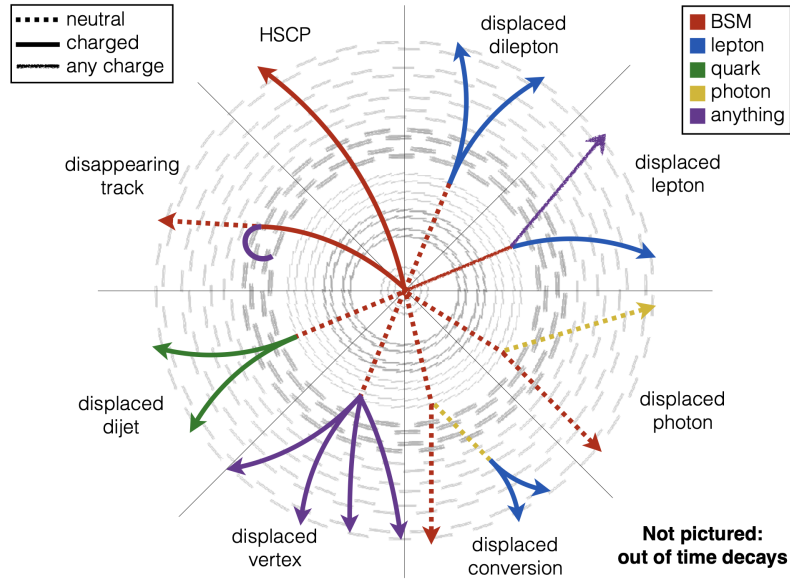


Figure 2.7: Schematic view of a variety of BSM LLP signatures in general-purpose detectors at the LHC shown in their transverse plane. Figure from Ref. [66].

2.3.2 Long-lived Heavy Neutral Leptons

The Seesaw mechanism does not place constraints on the mass and possible lifetime of HNLs, therefore it is feasible to extend the search range to HNLs manifesting

as long-lived signatures. In general-purpose detectors, the accessible lifetimes of long-lived HNLs are between $10 \mu\text{m}$ and 10 m , while the masses range between 1 and 20 GeV. No HNL signatures below 1 GeV are considered, as such HNL decays exceed the physical limits of the detector. They also typically proceed via two-body processes, where two quarks hadronise to form a meson. On the other hand, HNLs heavier than 20 GeV tend to decay promptly.

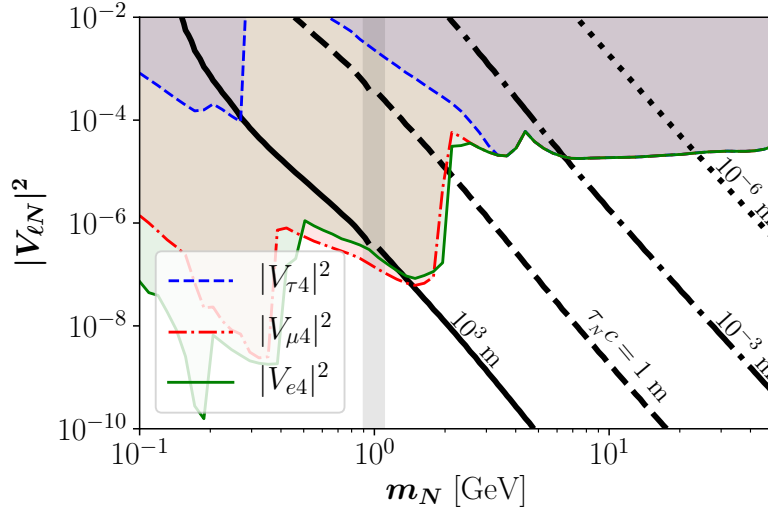


Figure 2.8: Contour lines for various HNL decay lengths, $c\tau_0$, assuming equal coupling to all three neutrino generations simultaneously. Shaded areas correspond to approximate current exclusion limits by different experiments. Figure from Ref. [67].

Direct searches for HNLs have been performed using model-independent phenomenological approaches, assuming that the HNL mass, M_N , and coupling to SM neutrinos of flavour ℓ , V_ℓ , are free parameters (see Ref. [59]). The lifetime (expressed as a distance in such searches), $c\tau_0$, of an HNL parametrically scales as a function of its mass and coupling strength to a single generation lepton ℓ , therefore it is a backward oblique line in the phase space diagram in Fig. 2.8. In a generic scenario where an HNL with a mass of a few GeV couples to all three generations simultaneously, the decay width Γ_N is approximately proportional to:

$$\Gamma_N \propto G_F^2 M_N^5 \sum_{\ell=e, \mu, \tau} |V_\ell|^2. \quad (2.64)$$

This equation is valid for both Dirac and Majorana-type HNLs.

Similar searches for long-lived HNLs relevant to the results of this Thesis have been previously performed by CMS [6], ATLAS [7, 68] and the LHCb [69] collaborations both via semi-leptonic and fully-leptonic decay channels. However, such searches

mainly focused on single-generation HNL couplings that are now disfavoured by theoretical predictions (see for example Fig. 2.5).

Chapter 3

Experimental Setup

This Chapter presents the experimental setup used to search for long-lived HNLs, as well as provides a brief introduction to future efforts in the field of HEP. The analysis presented in this Thesis uses data collected by the CMS experiment between 2016 and 2018 at the Large Hadron Collider, located at CERN. An introduction to the accelerator complex housed at CERN is presented, as well as the technology of the LHC. Furthermore, a detailed description of the various detector layers relevant to the long-lived HNL search is presented. Finally, a brief discussion focussing on the Level-1 trigger upgrade program of the CMS detector for the High-Luminosity LHC is provided.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the world's most powerful accelerator and one of the most complex machines ever built by humankind. The LHC is a circular synchrotron with two oppositely travelling beams of charged hadrons brought into collision at four interaction points (IP), each surrounded by a detector capturing the decay products of the collisions between the two beams [70].

The LHC was designed and constructed at the European Organisation for Nuclear Research (CERN) as a collaborative international project, committed to peaceful and open-data research in nuclear and high-energy physics. It is housed in a roughly 26.7 km circumference circular tunnel, 50-175 m under the Swiss-French border, near Geneva. The LHC tunnel was originally constructed for the Large Electron-Positron Collider (LEP), and was one of the most challenging civil engineering projects undertaken by humankind. LEP was used in the 1990s and was the largest lepton

collider to date, providing precision measurements of various SM particle properties (most importantly, the Z and W boson masses) and confirming the consistency of the SM, via its "clean environment" obtained by colliding elementary particles. Nevertheless, its main limitation for reaching higher collision energies resulted from the large energy loss arising from synchrotron radiation emitted by electrons:

$$\Delta E = \frac{q^2 E}{\epsilon_0 3R(m_0 c^2)^4} \quad (3.1)$$

where q is the charge of the accelerated particle to energy E on a circle of radius R , ϵ_0 is the vacuum permittivity, c is the speed of light and m_0 is the rest mass of the accelerated particle. As electrons are the lightest charged leptons of the SM, when accelerated to high energies in a circle, they emit much more synchrotron radiation than protons, due to their rest mass being orders of magnitude larger. The LHC exploits the fact that protons are composite particles where collisions take place between constituent partons, allowing it to access a broader energy range with less synchrotron radiation. However, hadron collisions result in a more complicated environment to study. The main limiting factor of hadron colliders is the magnetic field strength generated by the superconducting dipole magnets, discussed later in this Section.

The primary goal of the LHC is the discovery and study of the Higgs boson, by conducting experiments at the TeV scale and having the ability to thoroughly explore SM physics at the EWSB scale. Furthermore, part of the LHC's scientific program has been to conduct precision measurements of the SM and to study an exotic form of matter, the so-called *quark-gluon plasma*, by reconstructing the primordial state of the Universe after the Big Bang. In addition, LHC experiments have been searching for BSM phenomena (supersymmetry, ν MSM, Dark Matter etc.) aiming to confirm or disprove such hypotheses.

The LHC has multiple operational modes, primarily colliding proton-proton (p-p) beams, which is also the main focus of this Thesis; the descriptions below therefore focus on p-p data collection. Nevertheless, the LHC reserves a month each year for heavy ion collisions, including lead, oxygen and xenon.

Particles are brought into collision at four sites around the LHC ring, each IP being surrounded by an experiment, each with its own physics program. The CMS and ATLAS experiments are general-purpose detectors, that research a broad range of physics processes. They independently confirmed the discovery of the Higgs boson in 2012. The ALICE detector was designed to accurately test heavy ion collision environments in order to study the physics of the quark-gluon plasma. The LHCb

experiment has been designed to conduct precision measurements in the heavy flavour physics sector and to study the matter-antimatter asymmetry observed in the Universe. Of course, all the collaborations mentioned above have the ability to perform other types of physics searches outside their optimal design, providing a cross-check of the results from fellow LHC experiments.

Focusing on the p-p program of the LHC, Fig. 3.1 provides an overview of the CERN accelerator complex, bringing p-p beams to record energies. Protons are obtained from a canister of compressed hydrogen gas that is injected into the LINAC 2 linear accelerator, where an electrostatic field strips off the electrons from the hydrogen gas, leaving protons that are then accelerated to 50 MeV energy. Following that, the protons are transferred into a system of synchrotron pre-accelerators that gradually accelerate the proton beams before finally injecting them into the LHC ring: the proton synchrotron booster (PSB) accelerates the protons to 1.4 GeV, after which the proton synchrotron (PS) further accelerates them to 25 GeV; the protons are then passed into the Super Proton Synchrotron (SPS) which accelerates them to 450 GeV. The SPS then injects the protons into the LHC ring, which consists of 8 radio frequency (RF) cavities that accelerate the protons up to 6.5-7 TeV. As of December 2022, the largest proton beam energy achieved at the LHC was 6.8 TeV at the start of Run-3, colliding protons at record centre-of-mass energies of 13.6 TeV [71]. The data analysed in this Thesis was collected during Run-2, between 2016 and 2018, with $\sqrt{s} = 13$ TeV.

Proton beams at the LHC are organised in 2808 bunches of 1.15×10^{11} protons with 25 ns (~ 7.5 m) spacing, which are kept in circular motion by a Lorentz force created by an 8.4 T magnetic field. Such large magnetic fields around the LHC ring are generated using 1232 Niobium-Titanium superconducting dipole magnets, cooled to 1.9 K by the world's largest liquid nitrogen and helium cryogenic systems to date¹. The proton beams are brought into contact by focusing quadrupole magnets at a rate of 40 MHz. As mentioned before, the main limitations to the LHC reaching higher collision energies are the current technological capabilities of developing stronger superconducting magnets that can keep protons in circular motion and bring them into collision at IPs. In addition, higher-order magnets keep the proton beams focused in order to counteract the repelling Coulomb force between protons within a bunch, as well as at the IPs. In technical terms, p-p collisions are called "bunch crossings" and recorded interactions are referred to as "events".

The rate of event production in LHC collisions is given by:

¹Once completed, the International Thermonuclear Experimental Reactor (ITER) will become the "world's largest and most powerful fridge" [72, 73].

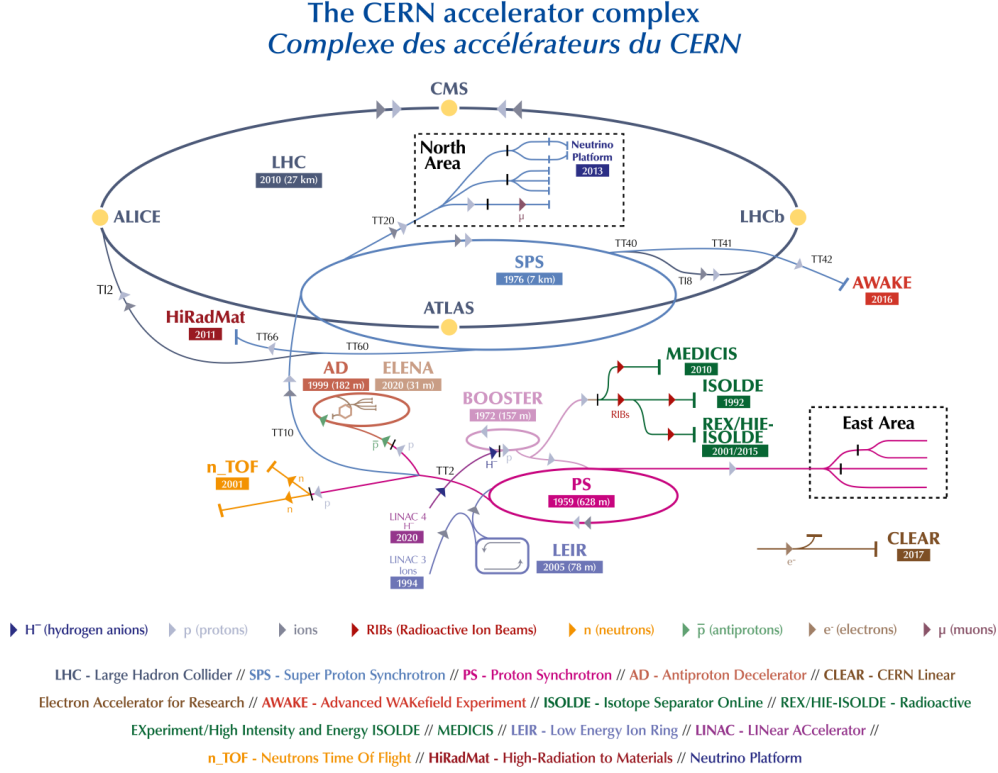


Figure 3.1: Schematic diagram of the CERN accelerator complex, highlighting various accelerator stages before the proton beams are injected into the LHC superconducting synchrotron ring. In addition, other particle extraction facilities and experiments are shown. Image credit: CERN.

$$\frac{dN}{dt} = \sigma L \quad (3.2)$$

where σ is the cross-section for a given process and L is the instantaneous luminosity determined by the particle beam parameters. The LHC design instantaneous luminosity was $1.0 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. However, during Run 2, the LHC achieved a record instantaneous luminosity of $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ with a further planned increase in Run 3. The integrated luminosity, $\mathcal{L}$, is given by the total number of recorded events collected by an experiment over its operation period, measured in inverse femtobarns $\text{fb}^{-1} = 10 \times 10^{-43} \text{ m}^2$:

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

Fig. 3.2 shows the total integrated offline luminosities recorded by the CMS detector during its operation periods, including Run 3, as well as the cumulative luminosity delivered by the LHC. During Run 2, CMS recorded 137 fb^{-1} of p-p collision data,

the analysis of which is described in Chapters 7-10 of this Thesis.

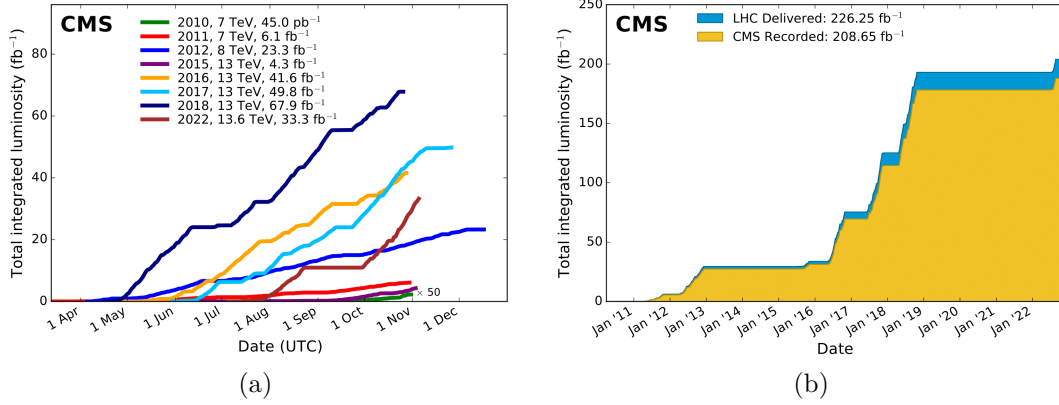


Figure 3.2: Diagrams showing the cumulative integrated luminosity versus day (left) and over the total operation period (right) recorded by CMS for p-p collisions at nominal centre-of-mass energy. The right diagram also shows the total integrated luminosity delivered by the LHC [74].

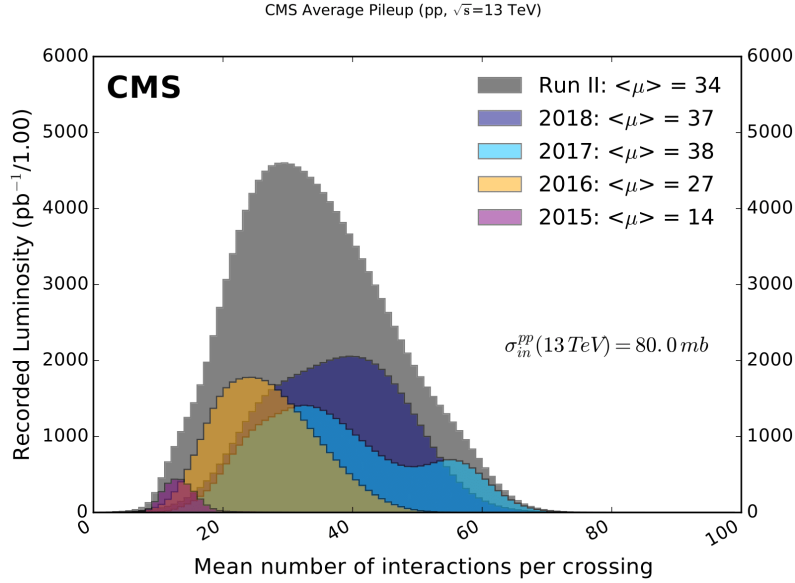


Figure 3.3: Distribution of the average number of interactions per bunch crossing (pileup) for p-p collisions during Run 2 (13 TeV) operation of the CMS detector. The overall mean values and the minimum bias cross-sections are shown [74].

In order to maximise the rate at which rare events occur in p-p bunch crossings, the geometry and number of protons stored in each bunch must be optimised. Protons that do not undergo collisions are kept in orbit and the bunch is brought into collision at the next IP. Over time, the LHC fill drops in luminosity due to protons being lost in collision events. The number of p-p interactions occurring per bunch crossing is called *pile-up* (PU). Figure 3.3 shows the PU distribution for the Run-2 operation of the CMS experiment, per year and in total. The two-peak feature in the 2017

pileup distribution is a result of the introduction of luminosity levelling in the CMS detector to avoid aggressive event filtering at the start of each LHC fill, such that the detector data storage system can keep up with the incoming event rate at higher luminosities.

3.2 The CMS detector

The Compact Muon Solenoid (CMS) experiment is one of the general purpose detectors at the LHC [75, 76]. The main physics goals governing its design were the discovery and precision measurements of the Higgs boson, as well as direct searches for BSM physics, such as supersymmetric (SUSY) particles, dark matter and ν MSM, among others. Furthermore, being a general-purpose detector, CMS is capable of making precision measurements for testing the SM and to search for signs of New Physics (NP).

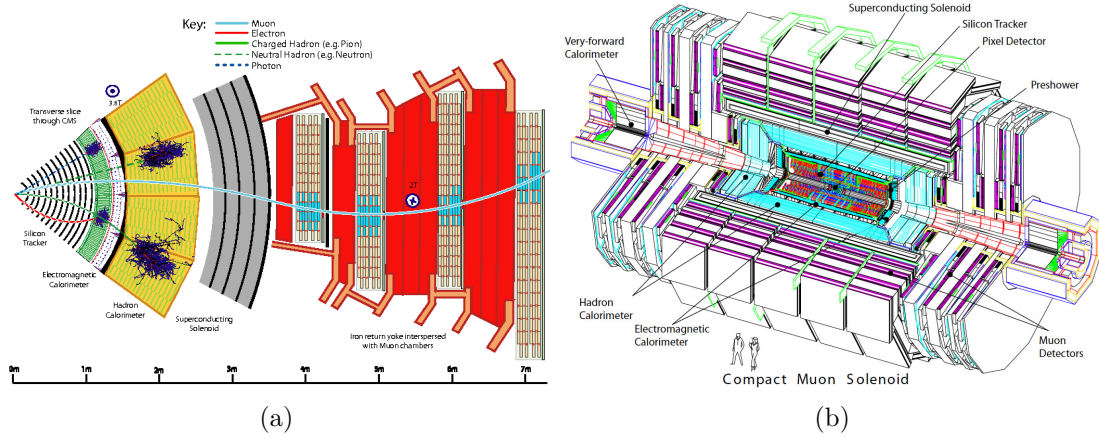


Figure 3.4: Exploded views of the CMS detector showing an onion-like layer structure and the effects of the strong magnetic field (2 and 3.8 T, respectively) on charged particles (diagrams from [77, 78]). The left figure is a section of the detector in its transverse view. The right figure is a 3-dimensional projection of CMS enhancing the concentric nature of the sub-detector layers (tracker, electromagnetic and hadronic calorimeters, superconducting solenoid and muon chambers, as well as the forward hadronic calorimeter) around the collision point in the centre of the beamline.

The CMS detector (shown in Fig. 3.4) is a spectrometer consisting of a hermetic, concentric multi-layer structure surrounding the LHC beam axis. Each detector layer serves a different purpose, with the cylindrical structure being divided into a *barrel section* and two disk-like *endcap sections* in the forward region. An important aspect driving the detector design and layout was the choice of the magnetic field configuration for precise charged particle (track) momentum measurements, as well

as charge identification under the Lorentz force generated by the magnetic field. Large bending power is needed to precisely measure the momentum of charged particles. Just outside the CMS calorimeters sits a 13 m long, 5.9 m inner diameter, superconducting solenoid yielding a magnetic field of 3.8 T at its centre [79].

The CMS detector consists of the following detector layers listed moving radially outward from the beam axis, which is embedded in a beryllium beampipe. The silicon tracking system consists of an inner layer of pixel detectors surrounded by a silicon strip tracking system. The role of the tracker is to accurately measure the trajectories (tracks) of charged particles moving outwards from the IP, as well as to locate the primary vertices of hard scattering processes and to identify potential secondary vertices. The tracker is surrounded by the electromagnetic calorimeter (ECAL) followed by the hadronic calorimeter (HCAL), each of which absorbs the energies of incoming particles, in order to measure the energy deposited. The ECAL focuses on the measurement of electrons and photons, whereas the HCAL absorbs hadronic particles that pass through the ECAL. The superconducting solenoid magnet was installed outside the HCAL and is surrounded by 4 layers of muon chambers embedded in the iron return yoke. The role of the return yoke is to guide and confine the magnetic field outside the solenoid magnet, besides providing structural support to the experiment. Fig. 3.4a presents a slice in the transverse plane of the CMS detector showing the response of the various detector layers to different particle types, as well as the typical trajectories of charged particles as a result of the strong magnetic field. The CMS detector weighs 14000 tons and has a length of 21.6 m and a diameter of 14.6 m. The experiment has some 75 million readout channels that are operated synchronously at 40 MHz bunch crossing frequency. The following subsections discuss the technical aspects of each detector system of the CMS experiment in more detail.

Geometry and coordinate system

For event reconstruction purposes, the CMS experiment adopts a right-handed Cartesian coordinate system with the point of origin at the IP. The z-axis is chosen to align with the beamline and points anticlockwise along the LHC ring. The y-axis points vertically upwards, while the x-axis points towards the centre of the LHC ring.

For simplicity and due to the cylindrical shape of the CMS detector, alternative coordinate systems are also used for various physics reconstruction purposes. These are based on polar coordinates: ϕ is the azimuthal angle in the transverse (x-y)

plane, relative to the x-axis; r is the radial distance; η denotes the pseudorapidity of a particle. Pseudorapidity is a transformation of the polar angle, θ , relative to the beam axis and the relativistic particle, and is defined as:

$$\eta = -\ln \tan \left(\frac{\theta}{2} \right). \quad (3.4)$$

For higher pseudorapidity values, particles travel closer to the beamline, consequently, they are considered to move in the "forward region" of the detector². For instance, particles emerging from PU interactions typically possess large η values. Another useful quantity is the angular separation between objects defined in the (η, ϕ) plane:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (3.5)$$

Certain physics processes of interest have a significant presence in the transverse plane of the detector, therefore it is useful to define the following quantities:

- transverse momentum of a particle: $p_T = \sqrt{p_x^2 + p_y^2}$;
- missing transverse momentum: $\vec{p}_T^{miss} = -\sum_i \vec{p}_T^i$ where $\vec{p}_T^i$ is the transverse momentum of each visible particle i in the detector;
- missing transverse energy: $E_T^{miss} = |\vec{p}_T^{miss}|$;
- transverse impact parameter, d_{xy} , which is defined as the shortest distance between a track and the beam line in the transverse plane;
- transverse impact parameter significance: $d_{xy}^{sig} = d_{xy}/\sigma(d_{xy})$, the fraction between transverse impact parameter and its uncertainty.

3.2.1 Tracking system

The tracking system of the CMS detector [80] is its innermost component, lying closest to the IP. The role of the tracker is to accurately reconstruct charged particle trajectories that are bent in the strong magnetic field created by the solenoid magnet, as well as to identify primary vertices from hard scattering processes in each collision.

²The forward region of CMS typically refers to the portion of the detector that is situated at smaller angles with respect to the beamline. It covers particles emitted at relatively small angles from the beamline in the direction of the beam.

In addition, the tracker has the ability to identify and locate secondary vertices from decaying particles, such as tau leptons or B hadrons.

Due to its proximity to the IP, the tracker faces the highest emerging particle flux, ($\mathcal{O}(1000)$) per bunch crossing. Consequently, the design of this detector system needs to prioritise robustness against radiation damage, as well as high spatial and temporal granularity and excellent resolution for precision tracking. Furthermore, the design of the inner tracker needs a low material budget to avoid unwanted scattering interactions that would degrade measurements of outgoing particles. These scattering processes include electron bremsstrahlung, photon conversion and hadron nuclear interactions.

To satisfy the requirements for radiation hardness, low material budget, fast and accurate response, high granularity, as well as excellent track and vertex resolution, the inner tracking system of the CMS experiment was built entirely of silicon detectors. The CMS tracker has a length of 5.8 m, a diameter of 2.5 m and pseudorapidity coverage of $|\eta| < 2.5$, and represents the largest silicon tracker ever built for a physics experiment. The fundamental working principle of silicon trackers is based on ionising radiation of passing charged particles in their depletion zone free of charge carriers. The ionising radiation creates electron-hole pairs in the silicon that are converted into an analogue signal, measuring the hits of the passing charged particle in each silicon detector layer. The momentum resolution of the tracker is 1-2% for outgoing particles with p_T up to 100 GeV in the central rapidity region. At higher energies, the resolution quickly deteriorates as the curvature of particle tracks decreases.

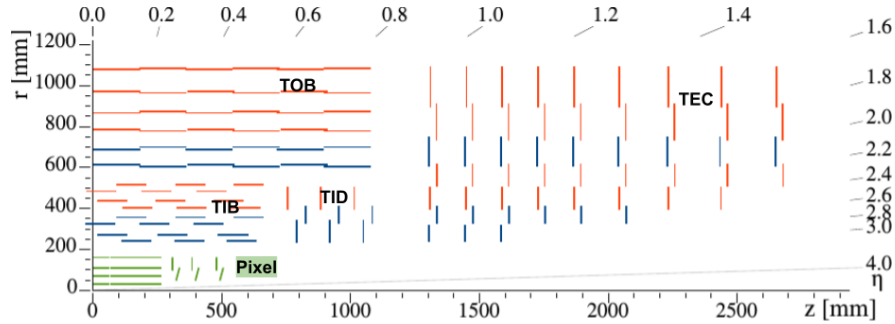


Figure 3.5: Schematic diagram of one-quarter of the Phase-1 CMS tracking system in r - z view. The pixel detector is closest to the interaction point and is shown in green. The silicon strip detectors are depicted as red and blue segments depending on whether they are single-sided or double-sided strip modules, respectively. Figure adapted from [81].

The tracking system of the CMS experiment consists of two main detector subsystems, both made of silicon, as shown in Fig. 3.5. The innermost layer is a pixel detector, extending to a radius of 20 cm around the beamline, receiving the largest particle

flux of all detector layers in the experiment. It has a symmetric cylindrical geometry around the beamline with the capability of measuring hit positions in 3 dimensions. Between 2008 and 2016, the pixel detector operated with 3 layers in the barrel region and 2 endcap disks in order to withstand a luminosity of $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, the default design of the LHC. After 2017, the luminosity delivered by the LHC was doubled and the CMS pixel detector was, therefore upgraded at the end of data taking in 2016. The Phase-I (2017-2025) pixel detector consists of 4 layers in the barrel region and 3 endcap disks, resulting in a total of 124 million individual pixels of size $100 \mu\text{m} \times 150 \mu\text{m}$. The transverse resolution of the pixel detector is $10 \mu\text{m}$, whereas the longitudinal one is $20 \mu\text{m}$. The total area covered by the pixel tracker is 1.9 m^2 with a $|\eta| < 2.5$ coverage. In addition to a more refined granularity and enhanced radiation hardness, the upgraded pixel detector has better track reconstruction efficiency and resolution performance compared to its predecessor [82].

The pixel detector is surrounded by several layers of silicon micro-strip detectors ranging in distance from 20 cm up to 110 cm radially around the beamline. They cover a total surface area of approximately 200 m^2 and $|\eta| < 2.5$, with 9.3 million strips operating under a reduced particle flux relative to the pixel detector, hence the use of less expensive silicon strip detectors. While some of these silicon strips have a tilted module configuration aiming to provide 2-dimensional hit positions and better coverage, most of the strips only provide 1-dimensional measurements. The silicon strip detector consists of four main subsystems. Between radii of 20-55 cm, one finds the Track Inner Barrel (TIB) and Track Inner Disks (TID), consisting of 4 barrel layers and 3 disks, respectively, with a resolution of $13\text{-}38 \mu\text{m}$ in the transverse plane. These are surrounded by the Tracker Outer Barrel (TOB) at 55-110 cm radii consisting of 6 larger strip layers placed in parallel with the beamline. To either side of the TOB, perpendicular to the beamline are the Tracker End Caps (TEC), which consist of 9 disks of micro-strips with similar dimensions to those in the TOB. The TECs and the TOB have a resolution of $18\text{-}47 \mu\text{m}$ [82].

Information from the tracker is crucial for most physics analyses to identify various particle types and their corresponding charge based on the track curvature. Nevertheless, signals from the tracking system are also used in the online trigger system to decide whether to keep events with interesting experimental signatures, as will be discussed in Sec. 3.2.6.

Tracking information has a crucial role in the HNL analysis presented in this Thesis, as it is needed for the correct identification of both prompt and displaced signature tracks associated with long-lived HNL decays.

3.2.2 Electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) [83] is a high granularity, compact, homogeneous detector with the role of efficiently measuring the energy and position of incoming electrons and photons that produce electromagnetic showers in the material of the calorimeter. High energy electrons emit bremsstrahlung radiation when penetrating dense materials whereas photons interact via electron-positron pair production with the same material.

The CMS ECAL is made of lead tungstate (PbWO_4) crystals, which have rather advantageous properties for effective energy measurements of photons and electrons. Lead tungstate has a short radiation length of $X_0 = 0.89$ cm, so electromagnetic showers have a limited longitudinal extent. In addition, it has a narrow Molière radius of only $r_M = 2.19$ cm, limiting electromagnetic showers in the lateral direction, as well. Besides being a radiation-hard material, lead tungstate has a fast response time³ of 25 ns, being in synchrony with the bunch crossing frequency [83].

The ECAL has two main detector subsystems, as presented in Fig. 3.6. The electromagnetic barrel (EB) consists of 61200 crystals with a depth of about $26X_0$, covering a pseudorapidity range of $|\eta| < 1.479$. Each crystal has a 0.0174×0.0174 coverage in the $\eta\phi$ plane, giving them a rectangular wedge shape. The advantage of this geometry is that no particle trajectory is completely aligned with the cracks between crystals. Therefore, there is relatively little loss of electromagnetic shower components in the ECAL crystals. The front face of the barrel crystals has a size of 22×22 mm, roughly corresponding to the material's Molière radius.

On either side of the EB, perpendicular to the beam axis lie the electromagnetic endcaps (EE), which are made of 7324 crystals with a thickness of about $25X_0$ and $1.479 < |\eta| < 3.0$ coverage. The front face of these crystals has a size of 28.6×28.6 mm and they are arranged in 5×5 crystal units. Just in front of the EE lie the electromagnetic pre-shower detectors (ES), which are sampling detectors with alternating layers of lead absorbers and silicon strips. The role of the ES is to distinguish displaced photons originating from neutral pion decays – originating from detector material effects – from prompt photons produced at the IP. The ES has a coverage of $1.653 < |\eta| < 2.6$.

When electrons and photons pass through the lead tungstate crystals, they emit a blue-green scintillation light that is captured by photodetectors placed at the rear side of the crystal blocks. Since lead tungstate has a relatively low light yield,

³80% of the emitted light collected in this time.

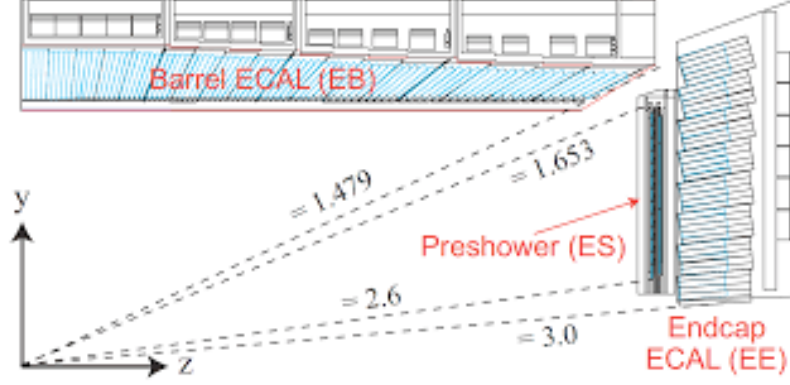


Figure 3.6: Schematic diagram of one-quarter of the CMS ECAL in r-z view, presenting the layout of the lead tungstate crystals in the electromagnetic barrel (EB) and endcap detectors (EE), as well as the position of the pre-shower detector in front of the EE. Figure taken from [84].

the scintillation light signal needs to be amplified before it is digitised. In the barrel area silicon avalanche photodiodes (APD) with an amplification factor of 50 are used, whereas the endcap has vacuum phototriodes (VPT) operating with an amplification factor of 10. The resulting optical signals are digitised by analogue-to-digital converters (ADC) at a 40 MHz rate. This set of discrete amplitude measurements is passed to the buffer waiting for the Level-1 trigger system to decide whether the collision event should be stored offline for later processing, as will be discussed in Sec. 3.2.6.

The energy resolution of the ECAL has been determined from electron test beam measurements [85]:

$$\left(\frac{\sigma}{E[\text{GeV}]}\right)^2 = \left(\frac{A}{\sqrt{E[\text{GeV}]}}\right)^2 + \left(\frac{B}{E[\text{GeV}]}\right)^2 + (C)^2, \quad (3.6)$$

where $A = 2.8\%$ is the stochastic term determined by event-to-event fluctuations of energy deposits; $B = 12\%$ is the noise term characterised by preamplifier and pile-up effects and $C = 0.3\%$ is the constant term assigned to account for intercalibration errors and crystal non-uniformity.

The ECAL has a crucial role in the analysis presented in this Thesis for the correct identification of electrons associated with long-lived HNL decays, as well as measuring the properties of the displaced jet. In addition, signals from the ECAL are employed in the study of electron identification scale factors presented in Chapter 9.

3.2.3 Hadronic calorimeter

The CMS hadronic calorimeter (HCAL) [86] is a sampling detector designed to measure the properties of collimated particles forming jets, by providing good hermiticity and containment. The HCAL is designed to measure the energy and position of not only charged hadronic particles originating from p-p collisions, but also neutral hadrons that do not leave tracks in the tracking system and interact only minimally with the ECAL. Due to its compactness and hermetic coverage, the HCAL is capable of measuring any energy imbalance in the transverse section of the detector arising from neutrinos or potential neutral NP candidates.

The HCAL consists of alternating layers of non-magnetic brass absorber plates interleaved with active layers of plastic scintillator tiles. When a hadronic particle arrives at the HCAL, it undergoes nuclear interactions with the dense absorber material, generating hadronic showers of particles detected via scintillation light in the active layers of the detector. The scintillation light is produced in the blue-violet region of the visible light spectrum and is subsequently converted by optical wavelength-shifting fibers to the green region. The signal is then amplified by a factor of 2000 by hybrid photodiodes to determine the energy deposited by incident hadrons.

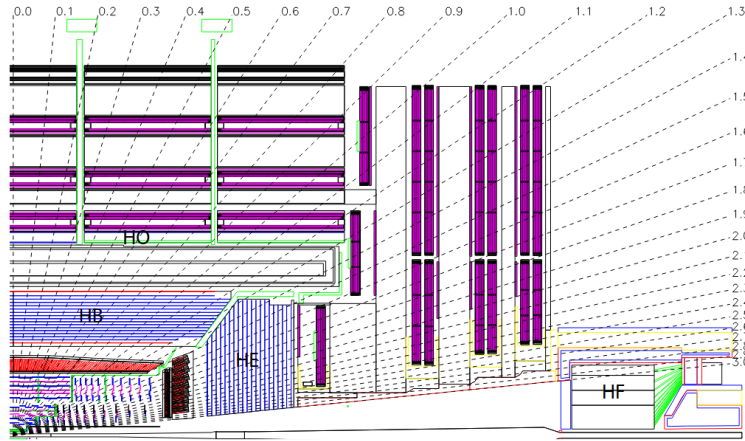


Figure 3.7: Schematic diagram of one-quarter of the CMS HCAL in r-z view, presenting the layout of the sampling detectors in the hadronic barrel (HB), endcap detectors (HE) and outer (HO) detectors, as well as the position of the forward hadronic (HF) detector down the beamline. Figure taken from [75].

The HCAL has 4 main components, as depicted in Fig. 3.7. The hadronic barrel (HB) covers a pseudorapidity range of $|\eta| < 1.3$ and consists of 36 identical azimuthal wedges split into 2 barrels. Each wedge has a spatial granularity in the (η, ϕ) plane of 0.087×0.087 , similar to the ECAL towers. The size of the HB is spatially constrained

by the position of the solenoid and the outer edge of the ECAL, lying radially from the beamline between $r = 1.77$ m and $r = 2.95$ m. This corresponds to 5.8 to 10.6 nuclear interaction lengths (λ_I) at increasing η values [86].

On either side of the HB, perpendicular to the beamline, lies the hadronic endcaps (HE), covering a pseudorapidity range of $1.3 < |\eta| < 3.0$. Up to $|\eta| < 1.6$, its spatial granularity corresponds to that of the HB, and beyond that it doubles in size, while maintaining a nuclear interaction length about $\sim 10\lambda_I$.

In order to mitigate the limited size of the HB, an outer hadronic calorimeter (HO) has been placed outside the solenoid magnet as a "tail catcher" for hadronic showers reaching $11.8\lambda_I$ depth and covering $|\eta| < 1.26$.

The HCAL is complemented by additional forward calorimeters (HF) placed ± 11.2 m from the IP along the beamline. The HF is designed to have extremely robust radiation hardness, since it receives high levels of particle flux at $3.0 < |\eta| < 5.2$. The HF is composed of thick steel absorbers embedded with quartz fibers that measure the energy deposited by forward jets via Cherenkov light produced by the showering particles.

The energy resolution of hadrons passing through the calorimeter system varies non-linearly, as well as the fraction of their energy deposit in the two calorimeter system [87]. Consequently, using cosmic ray muons and LHC beam data, the hadronic energy resolution of the barrel HCAL and ECAL combination are parameterised as [88]:

$$\left(\frac{\sigma_H}{E[\text{GeV}]} \right)^2 = \left(\frac{84.6\%}{\sqrt{E[\text{GeV}]}} \right)^2 + (7.4\%)^2. \quad (3.7)$$

As with the ECAL, signals from the HCAL are stored in buffers, and held until the Level-1 trigger decision is made on whether to retrieve the collision event.

Alongside the other detector layers discussed in this chapter, the HCAL is one of the key elements in detecting displaced jets, which are important for identifying the long-lived HNL signatures studied in the analysis discussed in this Thesis. In addition, the HCAL aids in the measurement of the missing transverse momentum, p_T^{miss} , of events used to train the binary event classifier presented in Chapter 8.

3.2.4 Solenoid magnet

The superconducting solenoid magnet of the CMS experiment [79] was built to aid the accurate measurement of the momentum and charge of charged particle trajectories bending under the Lorentz force it creates. With dimensions of 13 m in length and 5.9 m in inner diameter, it fits the tracking and calorimeter systems inside the cylindrical coil. The CMS magnet is capable of storing 2.6 GJ of energy when fully charged by a 19.14 kA current and creating its maximum 4 T magnetic field. It is the largest superconducting solenoid ever built for a physics experiment. However, due to durability considerations, the CMS magnet has actually been operated at 3.8 T. The solenoid magnet consists of a 4-layer winding of superconducting Niobium-Titanium (NbTi) fibers cooled to a temperature of 4.6 K.

The CMS solenoid magnet is surrounded by an iron return yoke system that is needed to contain and guide the magnetic field, as well as provide structural support to help the detector withstand the large forces created by the strong magnetic field. The steel yoke consists of 5 barrel wheels and 6 endcap disks interspersed with the muon chambers and weighs a total of 10000 tons (the largest contribution to the total mass of the CMS experiment). Due to its density, the iron yoke also serves to filter out any surviving hadronic activity from collisions, ensuring that only muons and weakly interacting particles, such as neutrinos can pass through them.

3.2.5 Muon system

The muon system is one of the most crucial detector layers of the CMS experiment [89], since the efficient, precise identification and measurement of muons has a critical impact in a multitude of SM and BSM searches. Muons are measured independently in three different detector layers of the CMS experiment (inner tracking system, after the solenoid and the muon system itself), yielding an impeccable muon reconstruction efficiency of 95-99% and a muon momentum resolution of $\sim 1\%$ ($\sim 10\%$) for muons with $p_T = 10 - 100$ GeV (1 TeV), demonstrating a robust performance for the muon system. Muons penetrate the calorimeter system without much interaction and energy loss, being only one of few particles surviving up to the muon chambers.

The muon spectrometer consists of 3 different detector technologies, all relying on gas ionisation technologies, as depicted in Fig. 3.8. The chambers are embedded in the iron return yoke, which also serves as an absorber to any surviving hadronic showers (known as "punch-through"). Beyond the yoke, only muons and undetected

neutrinos. The muons ionise the gas molecules in the chambers and the resulting electrons are collected, producing an electric signal.

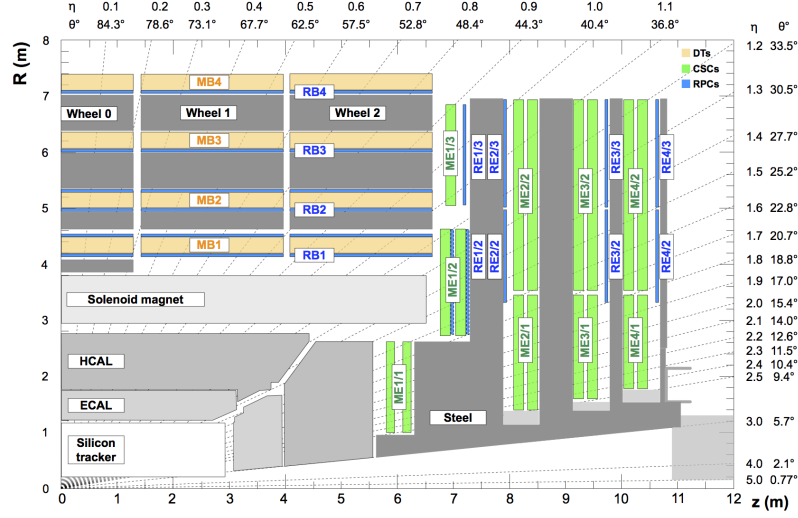


Figure 3.8: Schematic diagram of one-quarter of the CMS detector in r - z view, presenting the layout of all detector subsystems, but with a focus on the structure of the muon system. The various muon chambers are interleaved between the iron flux return yoke. The drift tubes (DTs) are located in the barrel region (labelled as "MB" in orange), while the cathode strip chambers (CSCs) are located in the endcaps of the muon system (labelled "ME" in green). The resistive plate chambers (RPCs) are located in both the barrel and endcap regions of the muon system (labelled "RB" and "RE" in blue). Figure taken from [90].

The barrel region ($|\eta| < 1.2$) of the spectrometer consists of drift tubes (DTs), which are suitable, as the muon rate, the neutron-induced background and the strength of the magnetic field are relatively low, with the latter also being fairly uniform. The endcap regions ($0.9 < |\eta| < 2.4$) of the muon system consist of cathode strip chambers (CSCs), where the non-uniform magnetic field increases, as well as the muon rate. These two detector subsystems provide excellent position reconstruction. Complementary to the DTs and CSCs, resistive plate chambers (RPCs) are added in both the barrel and endcap regions ($|\eta| < 1.6$), providing a fast response and good time resolution for the detected muons to be associated with the correct bunch crossing. However, RPCs have rather a coarse position resolution compared to. Information from all three subsystems is then passed to the Level-1 trigger.

The reconstructed hit spatial resolution of the muon system is about 50–300 μm , whereas the reconstructed hit efficiency ranges between 94–99%. The segment efficiency of the muon detector is about 97%. The resolution of muons with p_T up to ~ 100 GeV is 1% in the barrel and 3% in the endcap, while it reaches $\sim 6\%$ in the barrel for muons with $p_T = 1000$ GeV. The muon momentum resolution measurements were obtained using cosmic ray data and 13 TeV collision data, combining information

from both the inner tracking system and the muon detectors [90].

The muon system has an important role in detecting prompt and displaced muons associated with the HNL production and decays studied in this Thesis. The displaced muon identification and reconstruction systematic study presented in Chapter 9, relies on inputs from the muon system.

3.2.6 Data acquisition system

Collision event rates at the LHC reach an exceedingly large rate of 40 MHz, posing a major challenge for processing capabilities, bandwidth and data storage limitations while data taking is performed. The data associated with one collision event has a size of 1-2 MB, which means the 40 MHz event rate would correspond to a data production rate of 40 TB/s. The transfer rate for events sent to permanent storage is approximately 2 GB/s, 4 orders of magnitude lower.

To reduce these data rates to more manageable levels, CMS employs a two-tier online event trigger system run consisting of the Level-1 trigger (L1) followed by the High-Level trigger (HLT), each having a different design [77, 91–93]. The triggers significantly filter the incoming events by orders of magnitude while the detector collects data. Each trigger level performs fast online (partial) reconstruction of collision events by looking for pre-defined patterns, such as the presence of specific particles with particular properties, such as energy and p_T limits as well as isolation from other particles. These patterns are identified via detailed studies performed before individual LHC runs take place and trigger menus (i.e. set of selection criteria, known as *trigger seeds*) are created based on the interests of different analysis groups. Characteristic patterns for "interesting" processes (typically SM EW processes and possible BSM signatures) may contain specific objects such as leptons, photons, high- p_T tracks, energetic jets, high missing transverse energy carried away by undetected particles or energy sum quantities. Events that are disregarded by the trigger system are lost forever. "Uninteresting", well-known physical events are typically inelastic scattering of partons resulting in QCD multi-jet events that have high cross-section rates, also referred to as *pileup*. Filtering out such events keeps the incoming data rate under control: reducing online and offline event storage, lowering the bandwidth required to transfer data, optimising the storage space on disks, maintaining stable event rates and so on.

Some SM processes with high production rates may exceed bandwidth capacities, but which must be recorded for offline analysis. Such processes are filtered through

a so-called *prescaling* approach, which means that only a pre-set fraction of these events are actually recorded and sent for storage by the trigger system.

Level-1 trigger system

The Level-1 trigger (L1T) [94] consists of custom-built electronics that perform event reconstruction (data processing) using configurable field-programmable gate array (FPGA) chip devices with large data bandwidth and powerful parallel computation capabilities. The L1 trigger reduces the event rate from 40 MHz to 100 kHz, by a factor of 400. The L1T is a low resolution, but rather fast and flexible (through FPGAs), fixed-latency algorithm looking at simpler patterns (in particular for muons, electrons, photons, jets or τ candidates) using coarse readouts from front-end electronics from only the calorimeters and muon systems, creating Trigger Primitives (TP). The data is kept in buffers, located close to the detector. The total time allocated for making trigger decisions is $\lesssim 4\mu\text{s}$, corresponding to 144 bunch crossings. Hence, the time allocated to L1T calculations is less than $1\mu\text{s}$ as about $2\mu\text{s}$ are needed to transfer the signals between the detector and processors via high-speed optical links. If the trigger signal is delayed, the event is lost forever. Fig. 3.9 shows a schematic diagram of the Phase-I Upgrade of the CMS L1T used during Run 2 data taking. It consists of: primitive generators (top layer), trigger components (blue) and the Global Trigger (purple), which takes the final decision [95].

High-level trigger system

The second trigger stage is the High-Level Trigger (HLT) [97], a CPU-based PC farm with about 30,000 nodes, located on the surface, directly above the CMS detector. The software-based system further reduces the event rate to $\sim 1\text{ kHz}$ with a variable latency of about 200 ms on average [96]. The HLT is a more sophisticated software system, which employs the full granularity from all CMS sub-detectors and performs a quick event reconstruction, almost at offline-level quality. The HLT uses a more complex trigger menu than the L1T, taking advantage of the high-resolution objects it has access to, as well as full tracking information.

Once an event passes the HLT selection, it is sent to permanent storage and detailed reconstruction at the *Tier-0* Grid computing site. The CERN GRID computing infrastructure distributes data around the globe and uses parallel computing resources, as well as back-up systems from this network of computers. Despite the performant and optimised trigger system of the CMS experiment, a few PB of data are still collected yearly during data-taking periods.

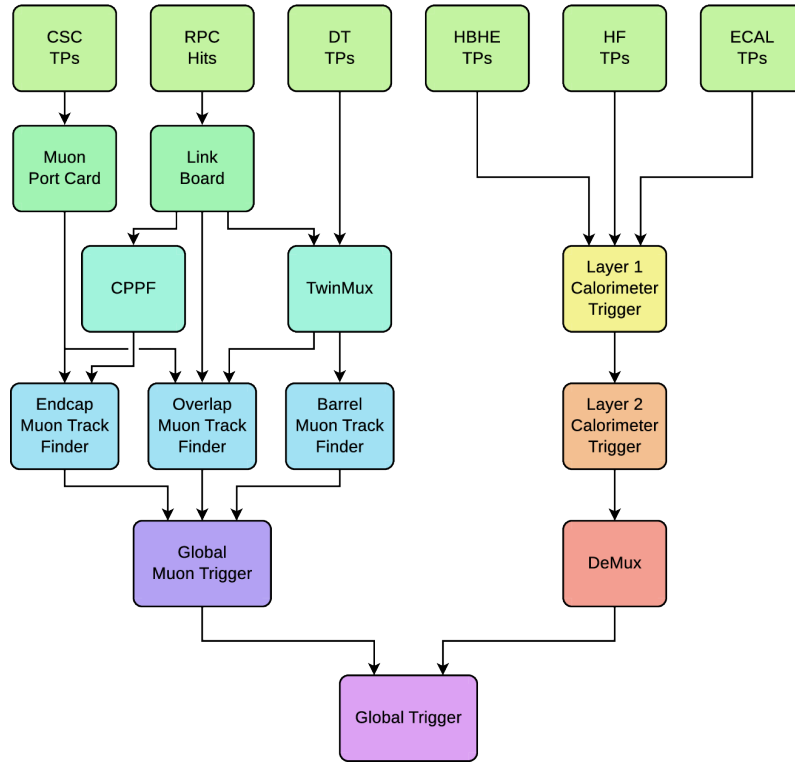


Figure 3.9: Schematic view of the upgraded Phase-I Level-1 trigger architecture used during the Run-2 data collection period of the CMS experiment [96].

3.3 The High-Luminosity LHC and CMS Phase-2 Upgrade

3.3.1 Overview

During its first three runs, the LHC is expected to collect 300 fb^{-1} of data. Since the discovery of the Higgs boson in 2012, no NP phenomena have been directly observed at the LHC, which has put the HEP community into a dilemma. Hypothetical, multi-TeV mass particles, as well as other exotic processes, may have such low cross-sections that they could be easily missed with the current LHC setup and trigger system. Moving towards precision measurements of well-understood SM processes, it will be necessary to increase data rates for improved event yield, particularly for unseen SM interactions with low production probability. Low cross-section processes of interest are the Higgs self-coupling in di-Higgs production, vector boson scattering processes or dark matter searches, among many others.

In order to address the low rate of "interesting" physics processes, an upgrade to higher luminosity has been approved and planned to be implemented around 2029. The High-Luminosity LHC (HL-LHC) would run at $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ instantaneous

luminosity on average, collecting 3000 fb^{-1} of data over 10 years of running. The highest instantaneous luminosity it is expected to reach will be $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. As a result of the increased luminosity, around 200 inelastic interactions would occur per bunch crossing [98]. This would increase the probability of rare interaction processes occurring, increasing the number of collected events. Nevertheless, the rate for low-interest pileup interactions will increase dramatically, as well. The great challenge of the HL-LHC will be controlling the incoming data rate: maximising the detection and recording of the desired rare processes while rejecting PU with optimal efficiency.

To address the high pileup conditions expected at the HL-LHC, extensive upgrade work will be conducted after concluding the Phase-1 LHC life cycle. The goal of such upgrade studies will be to reject pileup while maximising the acceptance of hard-scattering events of interest. This will be achieved by upgrading current detector technologies to higher granularity and improved radiation hardness, as well as improving the CMS trigger system. More details about the Phase-2 L1 trigger upgrade of the CMS experiment are given in Sec. 3.3.2.

3.3.2 Phase-2 Upgrade to the Level-1 trigger system

The HL-LHC will present a demanding environment with higher radiation and more challenging track and vertex reconstruction. Consequently, there will be a necessity for significant changes to both the LHC and the CMS detector. The Phase-2 upgrade of the CMS detector is scheduled to finish by 2028, which means that in the upcoming years, significant efforts will be invested in R&D and implementation work for the upgraded detector.

First of all, several layers of the current detector will deteriorate by the end of Phase-1 due to the high radiation deposited over the lifetime of the LHC. Radiation damage has a particularly pronounced effect on the tracker, since it lies closest to the IP point where the most energetic particles pass and deposit part of their energy, as discussed in Sec. 3.2.1. Furthermore, the scintillating materials of the calorimeters will gradually lose transparency when subjected to high radiation doses over time, especially in the endcaps. Hence, materials with better resilience to radiation damage and improved detection performance will need to be used.

To meet the demands of high PU conditions, increased granularity is required in all sub-detector layers, particularly for the upgraded tracker, as well as the calorimeter modules and muon system. High granularity will be needed to identify tracks and primary vertices associated with hard-scattering processes. With about 200

PU events on top of every hard scatter, this task will be rather challenging. The increased granularity of the tracker will ensure reduced channel occupancy, therefore hit merging in the pixels will be corrected to improve the distinguishability of spatially close tracks. Furthermore, a dedicated primary vertex (PV) finding algorithm will be added to the upgraded L1T. In order to reduce the effect of particle interactions with the tracker material, less material will be used to construct it. Finally, tracker information will be made available at the Level-1 trigger [99].

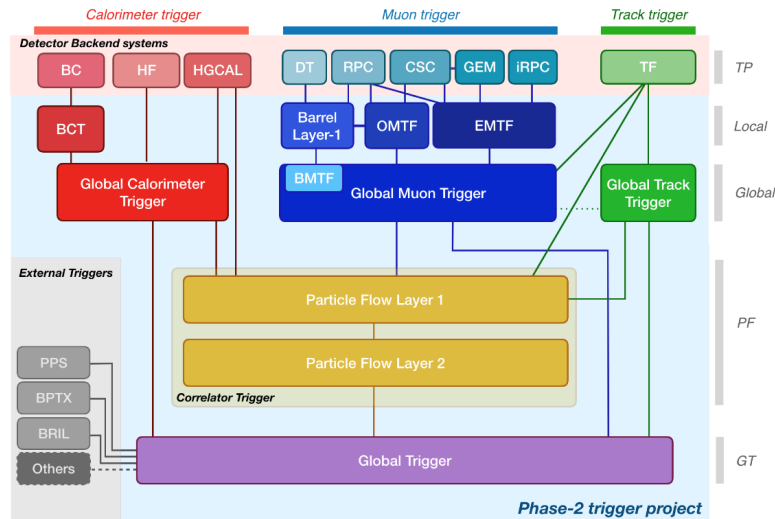


Figure 3.10: Functional diagram of the components to the Phase-2 Level-1 Trigger architecture [99].

As discussed in Sec. 3.2.6, the Phase-1 L1 trigger receives information from the front-end electronics of only three sub-detector layers: the two calorimeters and the muon chambers, but not the tracker. However, to mitigate high PU conditions at the HL-LHC, including the tracker information in the L1T will be a necessity in Phase-2. To make this possible, particle-flow-based reconstruction algorithms will be implemented on FPGA chips, with the potential of additional machine-learning-based software. Consequently, the L1T architecture will be restructured, allowing more sophisticated online reconstruction of physics objects and complex global event quantities, compared to its Phase-1 version. The latency of the trigger will be increased to $12.5 \mu\text{s}$, while the data rate passing the L1T will rise to 750 kHz. Preliminary studies show that employing Phase-1 L1T algorithms at HL-LHC PU 200 conditions will result in a data rate of 4000 kHz [100]. If the tracker information could be included, the rate would only be 500 kHz. Fig. 3.10 illustrates the proposed architecture for the Phase-2 L1T upgrade, encompassing the new sub-detector modules and increased granularity.

Chapter 4

Event Reconstruction

This Chapter describes the reconstruction and identification of essential physics objects used in the CMS experiment, with a particular focus on those employed in the long-lived HNL search presented in this Thesis. The chapter starts with, a brief discussion of the reconstruction of low-level event objects such as tracks, primary vertices and calorimeter clusters. Then, the discussion is extended to the methods used to combine these event objects to obtain physics objects describing the global event, such as muons, electrons, jets and missing energy. As the CMS detector was not explicitly designed to detect the displaced signatures manifested in the decays of long-lived HNLs, determining the reconstruction and identification performance for displaced physics objects is critical.

4.1 Particle Flow reconstruction

The offline reconstruction of event observables and physics objects is obtained via the particle flow (PF) algorithm [78], which combines various detector signals in an optimal manner to describe particles representative of the global event. Due to the fine granularity and hermeticity of the CMS detector, as well as the cutting-edge algorithms applied, the PF algorithm provides excellent measurements of kinematic particle properties and position, as well as highly efficient identification performance. Physics objects are first identified as PF candidates, such as leptons or charged/neutral hadrons, and eventually expanded into the reconstruction of more complex objects such as jets and missing energy.

4.1.1 Track reconstruction

Track reconstruction is an important step in the measurement of the momentum, charge, position and identification of charged particles. Charged particles interact with the tracker material, leaving hits in the pixel and strip detectors. In the magnetic field produced by the CMS solenoid, charged particles develop helical trajectories as they travel through the detector. The curvature of these helical tracks is determined by their momentum and charge, therefore reconstructed tracks provide crucial information on the kinematics of charged particles and can improve upon energy measurements delivered by the calorimeter system. Consequently, track reconstruction efficiency must be maximised.

Particle tracks are identified via a twelve-step iterative *combinatorial track finder* (CTF) algorithm [82]. Each iteration consists of four algorithmic steps: *seeding*, *trajectory finding*, *track finding* and *selection*. The first step of an iteration consists of selecting a trajectory seed based on pre-defined criteria on a handful of hits. Based on the track seed, a first rough estimate of the track parameters is provided. The second step extrapolates the trajectory of the particle based on the input parameters from the seed step, using a Kalman Filter¹ [102, 103]. The third step involves updating the track parameters and fitting hits along the estimated trajectory, adding them to the track. In the final step, tracks are selected based on pre-defined quality criteria such as the number of fitted hits layers, χ^2/ndof of the fit and whether they originate from a PV. This aims to reduce "fake" tracks from the final track collection. "Fake" tracks are misreconstructed tracks of poor quality with missing hits in at least one layer [78].

The first few iterations in the CTF algorithm aim to find tracks with high purity originating close to the IP. To reduce combinatorial complexity, hits assigned to tracks in one iteration are removed from subsequent iterations. In order to maximise track reconstruction efficiency, subsequent iterations relax quality criteria, with later iterations searching for high- p_T tracks and/or displaced tracks, relevant for the long-lived HNL search described in this thesis. Tracks found in all 12 steps of the iterative method are saved into the track collection of the p-p collision event [104].

¹The Kalman Filter is a "local fitter" since it estimates a joint probability distribution over the variables of interest by updating them following each new measurement. It permits the estimation of the evolution of a dynamic system, making the technique suitable for track building. The robustness of the method is the independence of multiple measurements on the same detector layer of which only one can lie on any track. The algorithm is named after Rudolf E. Kálmán, Hungarian-American engineer who was one of the primary pioneers of the theory in the 1960s [101]. The success of the algorithm is proven by the fact that it is widely used in HEP track fitting, as well as in other science and engineering applications.

The tracking efficiency is $> 99\%$ for isolated muons in the range of $1 < p_T < 100$ GeV, whereas for charged pions it is slightly lower ($\sim 90\%$) due to energy loss via nuclear interactions in the tracker material [82]. On the other hand, electrons lose much of their energy before even reaching the calorimeter system due to bremsstrahlung. Consequently, a dedicated tracking algorithm called the *Gaussian sum filter* (GSF) is used for electron track reconstruction, which takes into account such effects [105]. In addition, the GSF algorithm has the capability to identify electron-positron pairs resulting from photon conversions in the tracker material.

Track reconstruction is relevant to the studies of displaced lepton reconstruction and identification presented in Chapter 9.

4.1.2 Vertex reconstruction

The reconstruction and identification of primary vertices (PVs) close to the IP, as well as correct association of tracks with the reconstructed PVs is a crucial step in offline reconstruction of p-p collision events. As discussed in Sec. 3.1 in Fig. 3.3, around 34 p-p interactions were produced per bunch crossing during Run-2 in the CMS detector, however only about 25 PVs were successfully reconstructed per collision event on average, typically only one of them showing "interesting" physics features characteristic of a *hard-scattering* physics process. Tracks from other PVs are classified as pileup (PU) and are rejected.

To reconstruct a PV in a p-p collision event, tracks are clustered via a *deterministic annealing* (DA) algorithm [106] based on seed tracks consistent with their proximity to the IP. Candidate vertices with at least two tracks are fitted via an *adaptive vertex fitter* (AVF) in order to determine the vertex parameters, such as its coordinates and quality measures [107]. The hard scattering PV in a p-p event is chosen by the highest- p_T^2 sum of associated tracks, while the other reconstructed PVs in the event are discarded as PU [82].

Primary vertex reconstruction is relevant for the HNL search discussed in this thesis, particularly for the physics object reconstruction presented in the next Section 4.2.

4.1.3 Calorimeter cluster reconstruction

The energies deposited by incident particles via electromagnetic and hadronic showers are reconstructed in the two calorimeter systems of CMS. The shower deposits in the ECAL crystals and the HCAL components are clustered and a calibrated estimate

on the energy, position and timing of the cluster is provided. The calorimeter system is crucial in the reconstruction and measurement of photons and neutral hadrons, as they do not leave tracks [78, 108]. In fact, the calorimeter system provides additional measurements of the properties of charged particles that are complementary to those performed by the tracker. The PF algorithm combines information from the tracker and calorimeter systems to reconstruct charged particle candidates, as discussed in Section 4.2.

4.1.4 Muon tracking

The muon system delivers high purity measurements of muon tracks, as well as their correct identity (ID), as muons are the only charged particles that survive beyond the calorimeter system² [89, 90].

Muon seed tracks are first reconstructed using hit information from the DT and CSC systems via linear fits. Following that, a Kalman Filter-based track reconstruction algorithm extrapolates the muon track trajectories, using information from all three muon systems, including the RPC, as well. Muon tracks reconstructed using information only from the muon chambers are called *standalone muons*. The PF algorithm combines track information from the muon system with tracks reconstructed in the tracker to improve muon identification and kinematic reconstruction, as discussed in Sec. 4.2.1.

Muon track reconstruction is crucial for the derivation of systematic uncertainties associated with the reconstruction and identification of displaced muons, which is presented in Sec. 9.2.

4.2 Physics object reconstruction

4.2.1 Muon reconstruction

Two types of muon objects are reconstructed by the PF algorithm, the two being complementary to each other and maximising the offline muon reconstruction efficiency in CMS. As mentioned in the previous Section, tracks reconstructed in the muon chambers, called standalone (SA) muons, are combined with tracking information from the inner tracking system. By extrapolating SA muon tracks inwards to the

²punch-through will be discussed in Sec. 4.2.1

tracker, and associating charged tracks with them, one obtains *global muons*. Global muon tracks are fitted via a Kalman Filter algorithm, improving the reconstruction and momentum resolution of the muon tracks by including information from the two detector systems. This algorithm is more efficient for high energy muons, as they are more likely to reach the muon chambers and to leave several hits from which to reconstruct an SA muon segment. By extrapolating muon tracks reconstructed using tracker information with loose conditions applied ($p_T > 0.5$ GeV, $p > 2.5$ GeV) outwards to the muon chambers, while finding compatible hits in the muon system and taking into account detector effects such as the magnetic field, Coulomb scattering in the detector material and other energy losses, *tracker muons* are obtained. This algorithm is more effective for low-momentum muons (< 5 GeV) as it requires only at least one hit in the muon system to be identified as tracker muon. Muons reconstructed in the HNL analysis presented in this Thesis are required to be either global or tracker muons, about 99% of muons produced within the acceptance of the muon system being reconstructed as such [90]. Figure 4.1 presents a schematic diagram of the various muon types reconstructed by the PF algorithm and their reconstruction algorithms.

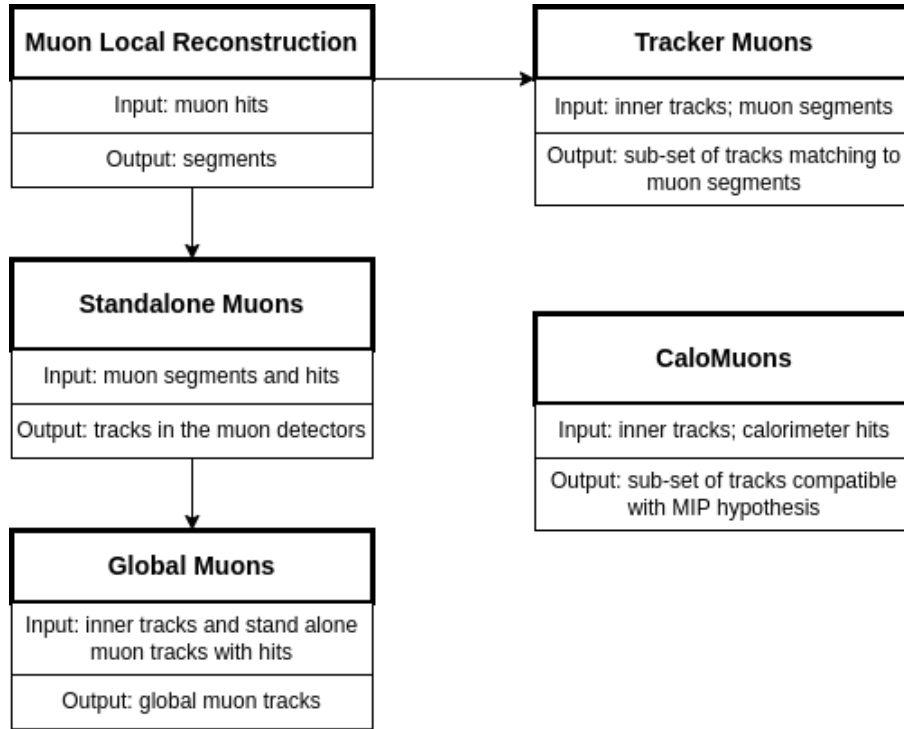


Figure 4.1: Flow-chart describing the different types of muons reconstructed within the PF algorithm in CMS. Segments are made of several hits. Global muons consist of standalone muon tracks reconstructed in the muon spectrometer and then extrapolated inwards to the tracker, finding compatible tracks. Tracker muons ensure that a track reconstructed in the tracker has matching hits in the muon station.

As mentioned previously, the CMS detector system is designed to ensure that only

muons reach the muon chambers from all SM particles except neutrinos that pass through undetected [89]. Nevertheless, occasionally, energetic hadronic particles, (usually associated with a jet) may reach and be detected by the first few layers of the muon spectrometer. This effect is called *punch-through* and is reduced by applying various quality criteria on the reconstructed muon tracks, as well as the identification of muon objects in the PF algorithm.

The HNL analysis presented in this Thesis uses two types of muon candidates. To capture prompt muons from W boson decays, that events are triggered on, *tight muon identification* criteria are required. Such prompt muons must be well-isolated from nearby hadronic activity resulting from PU. The relative isolation criterion is set to $I_{\text{rel.}}^\mu < 15\%$, where $I_{\text{rel.}}^\mu$ is defined as:

$$I_{\text{rel.}}^\mu = \frac{I_{\text{ch.had.}} + \max\left(0, I_{\text{neut.had.}} + I_\gamma - \beta \cdot I_{\text{PU}}\right)}{p_{\text{T}}^\mu}, \quad \beta = \frac{1}{2}, \quad (4.1)$$

where $I_{\text{ch.had.}}$, $I_{\text{neut.had.}}$, I_γ denote the summed energy deposits of charged hadrons, neutral hadrons, and photons within a cone of $\Delta R < 0.4$ around the muon direction, respectively. Contributions from PU tracks are taken into account, as well, by accounting for half their energy deposit as a correction to the neutral energy deposit. It is assumed that the pion isospin triplet (π^+ , π^- , π^0) has a production rate for each particle, therefore the neutral to charged track ratio in PU is expected to be half.

Tightly identified muons must be reconstructed as both a global and tracker muon. The inner tracker track must have at least six hits and at least one of them in the pixel detector. The global muon fit must be well-reconstructed tracks with $\chi^2/\text{ndof} < 10$ and have at least one hit in the muon stations. The tracker muon must have at least two matching segments in the muon chambers, avoiding fake muons from punch-through. Tight muons must originate close to the PV ($|d_{xy}| < 0.01$ cm; $|d_z| < 0.05$ cm), suppressing displaced muons resulting from the decays of b- and c-hadrons, as well as cosmic muons and PU tracks. Good efficiency is observed in the selection of tight muons of at least 90%, residual differences between simulation and data being corrected by (p_{T}, η) -dependent scale factors [90].

The other type of muon accepted in the analysis presented in this Thesis are displaced muons resulting from the decay of long-lived HNL candidates, which must pass *loose muon identification* criteria. Loose muons do not have to meet any isolation criteria or be compatible with originating from the PV of the event. In addition, they are not included in the event trigger criteria, therefore a looser transverse momentum threshold of $p_{\text{T}} > 3$ GeV is applied since the kinematics of the HNL decay products

are typically softer than those for prompt muons. The reconstruction and loose identification efficiency for displaced signal PF muons are presented in Fig. 4.2. Displaced PF muons are reconstructed with very high ($\sim 95\%$) efficiency for $p_T > 5$ GeV and transverse displacements below 10 cm. Highly displaced and low- p_T muons have suboptimal reconstruction performance due to the physical limitations of the tracker and its lower resolution for soft tracks, respectively. Nevertheless, the loose muon identification criteria show impeccable performance, close to 100%.

Studies of the modelling of displaced muon reconstruction and loose identification in simulation are discussed in Sec. 9.2, where (d_{xy}, p_T) and (d_{xy}^{sig}, p_T) -dependent scale factors are derived, respectively. These scale factors are incorporated in the systematic uncertainties of the final result of the analysis presented in this Thesis.

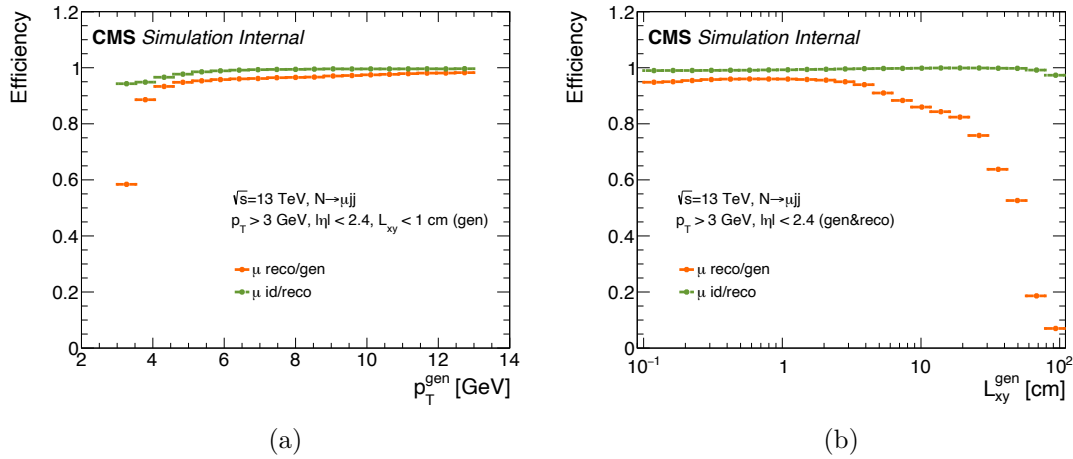


Figure 4.2: Signal muons from HNL decay reconstruction and identification efficiencies as a function of generator-matched lab-frame p_T and transverse displacement. Reconstruction efficiencies show a drop at low- p_T and with increasing L_{xy} . Figures from [109].

4.2.2 Electron reconstruction

Tracks linked to ECAL clusters can be reconstructed as electron objects by the PF algorithm [78]. As for the muon selection, tight and loose electron candidates are defined to target cases where either ℓ_1 or ℓ_2 is an electron. While the former uses an MVA approach in the identification process, the latter uses a *cut-based* approach relying on sequential selection [108]. Electron candidates found within $\Delta R < 0.05$ of a PF muon with $p_T > 3$ GeV are not selected to veto electrons radiated by a muon.

As with the muon selection, prompt electrons must be isolated, with their relative isolation being defined as:

$$I_{\text{rel.}}^e = \frac{I_{\text{ch.had.}} + \max\left(0, I_{\text{neut.had.}} + I_\gamma - \rho \cdot A_{\text{eff.}}\right)}{p_{\text{T}}^e}, \quad (4.2)$$

where the charged and neutral energy deposits within a cone of $\Delta R < 0.3$ around the electron candidate are used. The correction to the PU contribution $I_{\text{PU}} = A_{\text{eff.}} \cdot \rho$ is determined by the η -dependent effective areas and ρ denotes the median of the transverse energy density in $\delta\eta \times \delta\phi$ calculated from all PF particles within an event. Tight electron candidates in the barrel (endcap) region need to originate from the vicinity of the PV within $|d_{xy}| < 0.05$ (0.10) cm and $|d_z| < 0.10$ (0.20) cm. As mentioned before, tight electrons rely on MVA-based selection criteria based on the variables shown in Table 4.1, as well as other variables described in [108]. The MVA-based electron identification approach is more robust and has excellent performance in identifying prompt electrons from W boson decays, which events are triggered on [108].

Loose electron identification uses a custom cut-based selection, ensuring sufficiently high efficiency in finding displaced electrons resulting from the decay of a long-lived HNL. Table 4.1 shows the variables and the selection criteria applied for the cut-based selection, derived by relaxing certain requirements of the standard loose electron ID, with different criteria required in the EB and EE detector regions. Various shower-shape variables are used such as the ratio of hadronic to electromagnetic energy ($E_{\text{had.}}/E_{\text{EM}}$) and the second moment of the log-weighted distribution of crystal energies in η ($\sigma_{\text{in}\eta}^{5 \times 5}$). In addition, the electron track properties are exploited via the $|1/E_{\text{SC}} - 1/\rho|$ variable, where E_{SC} is the supercluster energy and ρ is the track momentum. Two angular variables ($|\eta_{\text{seed}} - \eta_{\text{track}}|$ and $|\phi_{\text{seed}} - \phi_{\text{track}}|$) are used, representing the angular separation of the seed cluster and the extrapolated track position. Displaced electrons arising from prompt photons interacting with the tracker material and consequently converting into an electron-positron pair are rejected if two such tracks can be fitted to a common vertex. Loose electrons must pass a p_{T} threshold of 5 GeV and stay within the tracker volume ($|\eta| < 2.4$). Figure 4.3 shows the reconstruction and loose identification efficiency for signal electrons in the analysis. Both the reconstruction and identification of displaced electrons have much poorer performance than muons due to the more complex signatures electrons leave in the detector, as well as muon reconstruction benefiting from the excellent performance of the muon spectrometer. The best electron reconstruction efficiencies are $\sim 80\%$ at $p_{\text{T}} > 10$ GeV and $L_{xy} < 0.6$ cm, dropping to $\sim 50\%$ at $p_{\text{T}} = 5$ GeV and $L_{xy} \approx 6$ cm.

The modelling efficiency of displaced electron loose identification in simulation is

Table 4.1: The applied custom loose cut-based electron identification criteria.

Quantity	ECAL barrel	ECAL endcap
$\sigma_{i\eta i\eta}^{5 \times 5}$	< 0.0112	< 0.0425
$ \eta_{\text{seed}} - \eta_{\text{track}} $	< 0.00377	< 0.00674
$ \phi_{\text{seed}} - \phi_{\text{track}} $	< 0.0884	< 0.169
$E_{\text{had.}}/E_{\text{EM}}$	$< 0.05 + 1.16/E_{\text{SC}} + 0.0324\rho/E_{\text{SC}}$	$< 0.0441 + 2.54/E_{\text{SC}} + 0.183\rho/E_{\text{SC}}$
$ 1/E_{\text{SC}} - 1/\rho $	< 0.193	< 0.111
Conversion veto	yes	yes

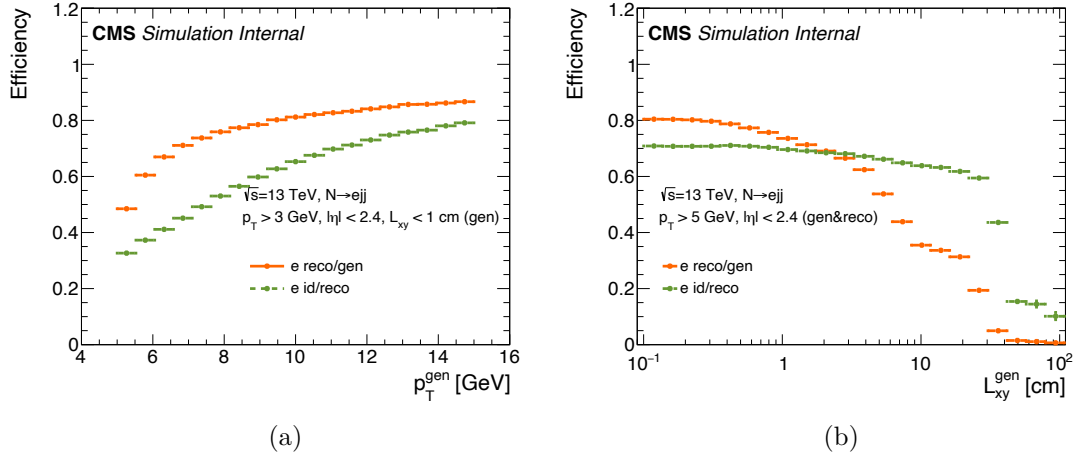


Figure 4.3: Signal electrons from HNL decay reconstruction and identification efficiencies as a function of generator-matched lab-frame p_T and transverse displacement. Electron reconstruction efficiencies show poorer performance compared to muons due to the more complex signatures they leave in the CMS detector system. Figures from [109].

further studied in Sec. 9.3, where $(d_{xy}^{\text{sig}}, p_T)$ -dependent scale factors are derived based on photon conversion studies. These scale factors contribute to the systematic uncertainties in the final result of the analysis presented in this Thesis.

4.2.3 Jet reconstruction

In p-p collisions at the LHC, proton-constituent partons interact with each other in the so-called hard-scatter, usually producing other final-state partons in the process. Nevertheless, quarks and gluons produced as part of a hard-scattering process typically hadronise into a collimated shower of hadronic particles known as *jets*. The properties of the parent parton that initiated any given jet can be determined by combining the information from reconstructed PF candidates forming the jet.

Jets in CMS are reconstructed from PF candidates clustered using the anti- k_T algorithm [110]. The jet momentum is determined as the vector sum of all particle momenta in the jet. The inputs to the anti- k_T algorithm are either PF candidates or previously clustered PF candidates. The anti- k_T algorithm relies on the following two variables:

$$d_{i,B} = \frac{1}{p_{T_i}^2} \quad (4.3)$$

$$d_{i,j} = \min\left(\frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2}\right) \frac{\Delta R_{i,j}^2}{R^2}, \quad (4.4)$$

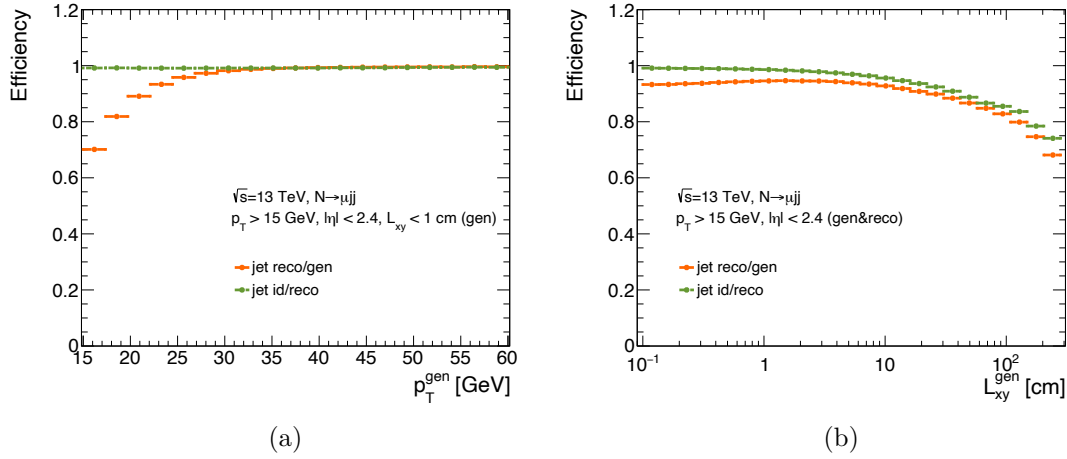
where $d_{i,B}$ is the distance between a PF particle cluster and the beam; $d_{i,j}$ is the distance between pairs of clusters i and j ; $\Delta R_{i,j}$ is the angular distance between clusters i and j and R is the benchmark cone size chosen to be $R = 0.4$. After calculating and ranking all distances $d_{i,B}$ and $d_{i,j}$ between pairs of clusters, if the smallest distance is of an inter-jet cluster, the two clusters i and j are merged into one and the algorithm recalculates all distances. However, if the smallest distance is of a beam-jet pair, the cluster i is considered the final jet state and removed from the jet collection, followed by all distances being recalculated. The iterations end when no PF candidates remain to be clustered into jets. As a result, this algorithmic procedure leads to mostly cone-shaped jets, with some exceptions when two hard particle candidates overlap within an angular distance of $2R$ from each other. In such a case, the algorithm must adjust the jet shape.

The influence of pileup is mitigated using the charged hadron subtraction (CHS) technique [111], where charged jet constituents not belonging to the PV are removed, making the reconstructed jet objects resilient against charged PU. The analysis described in this Thesis employs PF jets for offline reconstruction with improved energy and position resolutions compared to other types of reconstructed jets within the CMS experiment. The overall PF jet energy is mostly obtained from charged hadron tracks (around 65% of the total jet energy), as well as photons (contributing around 25% of the total jet energy measurement) and neutral hadrons ($\sim 10\%$ of total measured jet energy).

The reconstructed jet energy does not typically match the energy of the parton that initiated the jet, due to differing performance in the various subdetectors, losses in uninstrumented areas of the detector, and random fluctuations in the hadronisation process of partons. Therefore, the reconstructed jet must be calibrated using jet energy corrections (JECs) derived from the simulation of the detector response after comparison to the data [112, 113]. An offset correction is applied to the jet transverse

Table 4.2: Jet identification criteria for 2016 and 2017/2018 data-taking scenarios.

Variable	2016	2017/2018
Neutral hadron fraction	< 0.90	< 0.90
Neutral EM fraction	< 0.90	< 0.90
Number of constituents	> 1	> 1
Charged hadron fraction	> 0	> 0
Number of charged constituents	> 0	> 0
Charged EM fraction	< 0.99	-

Figure 4.4: Signal jets from HNL decay reconstruction and identification efficiencies as a function of generator-matched lab-frame p_T and transverse displacement. Figures from [109].

momentum to account for contributions from pileup. Further corrections are then applied to account for the non-uniform detector response as functions of η and p_T of the jets. In addition, residual corrections are applied, which account for jet p_T differences observed between data and simulation. Finally, jet energy smearing is applied to simulated jets to match the jet energy resolution (JER) observed in data. Each correction listed above has an associated statistical and systematic uncertainty propagated to the overall JEC values [114, 115].

The analysis considers central jets within $|\eta| < 2.4$ whose calibrated transverse momentum is greater than 20 GeV and which pass tight PF jet identification criteria displayed in Table 4.2, ensuring that spurious jets arising from detector noise are omitted while maintaining a high efficiency for genuine jets. Figure 4.4 shows the reconstruction and tight identification efficiency for signal jets originating from HNL decay in the analysis. PF jets are reconstructed with very high ($\sim 100\%$) efficiency for $p_T > 30$ GeV and transverse displacements below 10 cm ($\sim 95\%$). Highly displaced and lower- p_T jets have suboptimal reconstruction performance due to the finite jet energy resolution. Nevertheless, tight jet identification criteria show impeccable

performance, close to 100% for $p_T > 20$ GeV and $L_{xy} < 10$ cm.

Deep neural network-based displaced jet tagger

A novel deep neural network-based (DNN) displaced jet tagger is employed in the analysis presented in this Thesis with the aim of capturing displaced jets originating from the semi-leptonic decays of long-lived HNL particles. The DNN jet tagging architecture is based on the one presented in Ref. [9], which was the first CMS machine learning (ML) technical paper proving the feasibility of using domain adaptation (DA) in the context of HEP research.

The DNN jet tagger was developed to identify displaced jets manifesting different characteristics from SM jets, having the ultimate goal of finding jets originating from BSM processes. The baseline DNN architecture was inspired by recent b-tagging algorithms developed in the CMS experiment aiming to capture jets resulting from the decay of hadrons containing a b-quark [116]. As B-hadrons have a lifetime of $\mathcal{O}(10^{-12}\text{s})$, they travel a finite distance in the detector (1-10 mm) before decaying.

In contrast to b-tagging, searching for BSM jets has the challenge of not knowing the lifetime, hence the displacement of the tagged jets. As a result, the lifetime of the long-lived particle (LLP) is a free model parameter in the DNN architecture. BSM jets ranging from detector-prompt signatures to jets reaching the end of the inner tracker are considered by the displaced jet tagger. One advantage of the DNN tagger is the lack of an explicit requirement on reconstructed secondary vertices (SVs) in the collision event, which eventually extends the phase-space that can be accessed within the analysis. Moreover, the use of DA mitigates the need to calculate scale factors implemented in simulation for each input variable employed by the tagger (exploiting the order of 700 low-level event features) to account for any mismodelling issues between simulation and data.

A more detailed description of the DNN architecture applied in the long-lived HNL search, as well as its classification and performance, is provided in Sec. 7.5 and Ref. [109].

4.2.4 Missing transverse energy reconstruction

Since colliding protons and consequently, their constituent partons have negligible transverse momentum, due to energy-momentum conservation laws, the total momentum in the transverse plane after collision is expected to be zero. Thus, any

missing momentum or imbalanced energy in the transverse plane might indicate the presence of particles that do not interact with the detector material either as SM neutrinos or NP particles. Therefore, it is crucial for the CMS experiment to accurately reconstruct the total transverse momentum and energy of all PF particles. Missing transverse momentum and energy are physics variables used in physics analyses quantifying the imbalance of momentum and energy in the transverse plane of the detector, as defined in Sec. 3.2 [117].

However, the measurement of p_T^{miss} can be biased by detector effects such as inefficient tracking, minimum energy thresholds in the calorimeters, non-linearities in calorimeter response, electronics noise or the misreconstruction of muons as hadrons. In addition, misreconstruction of the p_T^{miss} variable might be caused by background effects such as cosmic rays or beam-halo particle collisions. Anomalous high- p_T^{miss} events resulting from spurious detector effects or physics background are discarded by dedicated algorithms and filters in order to avoid them being identified as potential BSM events. These filters have been applied in the HNL analysis as part of the pre-selection requirements for events.

Following the corrections applied on the reconstructed jet p_T in the event, the p_T^{miss} scale must be adjusted accordingly via similar jet energy corrections:

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

where the sum runs over all corrected jets with $p_T > 15$ GeV. Jets that deposit more than 90% of their energy in the ECAL are not included in the sum, ensuring that photon and electron jets are omitted in the p_T^{miss} correction process. The four-momenta of any muons found in jets are subtracted when performing the correction and added back to the corrected object [117].

The transverse energy is one of the most essential input variables in the binary event classifier separating signal HNL events from SM background, as discussed in Chapter 8.

Chapter 5

Analysis techniques

This Chapter describes the statistical analysis techniques employed throughout this Thesis. The first part of the Chapter provides a brief introduction to the Monte Carlo event simulation method, employed to model various key physics observables in the collected p-p data, as well as to help with event selection optimisation and training multivariate algorithms (MVA). The second part of this Chapter gives a brief description of the working concepts of the two main MVA algorithms adopted in the search for long-lived HNL particles: *boosted decision trees* (BDTs) and *neural networks* (NNs). Finally, various statistical formulations are presented, related to uncertainty calculations and statistical inference, which are employed in evaluating the results of the analysis via a frequentist maximum likelihood framework.

5.1 Monte Carlo event generation

To compare theoretical predictions of particle physics events with the reconstructed data from the detector, multi-step Monte Carlo (MC) generators are used to translate parton-level physics interactions into particle-level events, which are then passed through a detector simulation tool [118]. In addition, MC simulated events help understand the modelling of data, the main background sources and validate control over them, help in defining event selection and categorisation, in training Multivariate Analysis (MVA) discriminants, as well as performing statistical inference. The MC event generation consists of four steps detailed below.

Parton-level event generation

First, *parton-level* hard-scattering processes are generated by modelling the initial-state distribution of partons inside the colliding protons and computing the relevant matrix element (ME). Hard-scattering processes of interest can be both SM or BSM physics processes. The MC method samples such hard-scattering events from the corresponding cross-section integral predicted by theory. The advantage of the MC method is the ability to simulate high-dimensional final-state parameter space integrals where the uncertainty is independent of the integral's dimensionality but only depends on the number of simulated events (N) by $\sim \sqrt{N}$. Therefore, it is crucial to generate simulation sets with high event yield, also called *high statistics* simulation sets.

The *parton shower* (PS) simulates the emission of soft gluons and quarks from the initial and final-state partons. This process involves the successive branching of partons into lower-energy partons following random stochastic processes. The final-state partons eventually hadronise into jets of particles. Particles (gluons, quarks, photons etc.) emitted from the hard-scattering partons involved in the proton-proton collision are called *initial state radiation* (ISR), whereas particles emitted from particles produced in the hard-scattering process are called *final state radiation* (FSR).

Hard-scattering processes are generated at finite precision, in terms of contributing partons to the main diagrams of the processes. Leading order (LO) processes include only tree-level diagrams in the simulated physics process, whereas higher order (next-to-leading order, NLO or next-to-next-to-leading order, NNLO) diagrams include additional loops, providing higher precision to the simulations but also considerably more computational time, as well. The level of precision, as well as the MC event generator tools used in the simulation of background hard-scattering processes employed in the analysis presented in this Thesis are stated in Sec. 7.2. The MADGRAPH event-generator is used to simulate signal HNL events at LO precision studied in the analysis [119].

The ME simulation is dependent on the power of α_s which is calculated to a finite order in the simulation process. The associated radiative QCD emissions follow perturbative QCD models, typically of higher energies than those in PS. The ME calculation of soft gluon emissions in the simulated hard-scattering processes may generate inaccurate results, whereas the PS lacks strength in accurately modelling hard gluon emissions from initial- and final-state partons. Including radiative emissions from both ME and PS results in an overestimation of hadronic activity

in simulated events. To avoid this double-counting issue, various jet matching or merging schemes are employed [120, 121].

Events from a certain process are generated with equal probability to contribute to the cross-section integral. An event weight can be assigned to every MC generated event, based on the associated process cross-section, σ , and the number of generated events, N , to ensure the simulated sample represents the integrated luminosity of the collected data set, $\mathcal{L}$:

$$w_{\text{gen}} = \frac{\sigma \mathcal{L}}{N}. \quad (5.1)$$

If few events are generated, the event weight for an individual event will be large. For instance, the simulation of QCD multijet events is extremely computationally demanding with large theoretical uncertainties. As few such events are generated, the associated generator-level event weights are orders of magnitude larger than those for simulated events associated with other SM processes. Such events manifest themselves as sharp peaks with large uncertainties in MC simulation to data comparison histograms of kinematic observables due to their statistically limited yields with large generator event weights, as seen in the plots presented in Sec. 7.4.3. In addition, simulated samples with parton showers simulated at NLO precision may give rise to (unphysical) negative event weights to account for interference, singular phase space and particle emission (double) counting corrections [122]. At an integrated sample level, these negatively weighted events should cancel out with the majority of positively weighted events. MC simulated events with negative generator weights become problematic in phase-space regions with low expected event yields, their integrated sample yield possibly going negative, which is unphysical. Moreover, MC events with negative generator weights are an issue when training MVA classifiers, as their interpretation is misleading for the algorithm (as discussed in Sec. 8.3).

Particle-level event simulation

The next step in the MC event generation chain is the conversion of parton-level events into *particle-level* shower history. Final-state coloured partons from hard-scattering processes and PS go through a hadronisation process that binds them into colour-neutral baryonic and mesonic states. At this stage, any unstable particles in the event will ultimately into detector-stable particles, such as electrons, muons, photons, neutrinos, protons, neutrons, or mesons such as pions and kaons.

At this energy scale ($\sim$ few GeV), the QCD strong coupling constant becomes $\sim \mathcal{O}(1)$

and perturbation theory breaks down. In this non-perturbative regime, dedicated empirical hadronisation models are used such as string fragmentation and cluster models [120, 121].

Simulation of the underlying event

Secondary interactions unrelated to the hard-scattering process must also be simulated, such as multiple parton interactions, PU, and beam-beam remnants, called *underlying event* (UE) [123]. In the MC simulation sets employed in the analysis described in this Thesis, PU interactions are overlaid with the hard-scatter events of interest. Both the parton-level and particle-level interactions in the PS and the UE are generated with the PYTHIA event-generator [124]. The PDFs used by the event-generator to model the distribution of initial-state particles in colliding protons are NNPDF3.0 LO and NLO [125, 126].

Detector simulation

In the final stage of the simulation procedure, final-state particle-level events are passed through a simulation of the CMS detector based on the GEANT4 simulation toolkit [127]. This simulation package propagates final-state particle interactions with the CMS detector materials producing secondary particles, as well as simulating their dynamics in the magnetic field present throughout the CMS detector subsystems. In addition, the GEANT4 toolkit simulates the electronic read-out channels via a digitisation step. Consequently, similar event reconstruction techniques can be applied to simulated particle physics events as in real data.

Details of the signal and background simulation sets employed in the long-lived HNL analysis are given in Secs. 7.1 and 7.2.

5.2 Machine learning

Typical problems in HEP research involve the detection of very small signal yields over an order of magnitude larger background sample. Therefore, sophisticated statistical techniques have been developed to improve signal-over-background (S/B) ratios in selected phase spaces, typically applying a series of event selection requirements, also called physics observable "cuts". Following extensive studies on the modelling of the observables of interest, the optimisation of selection requirements to maximise (S/B)

ratios was typically done by hand. As analyses became increasingly more complicated, particularly in hadronic collisions with increased background processes compared to lepton colliders, manual tuning of variables became increasingly involved and time-consuming. The advent of machine learning (ML) applications in HEP provided an exciting opportunity to discover rare physics processes over large background samples, while the tuning of the event selection is performed in an automated manner by the ML algorithms.

Typical HEP problems consist of classification processes identifying signal processes, while simultaneously rejecting background in large data sets. Hence, the binary class system would be signal S versus background B . Another type of ML-related problem occurring both in physics and other disciplines is regression tasks, trying to predict a single continuous value mapping the inputs into a continuous function. Typical regression tasks in HEP relate to object reconstruction and identification. The application of ML techniques in HEP provided a two-way collaboration opportunity between physics research and the field of ML algorithm development, leading to state-of-the-art results in both research areas. The following Section provides a brief introduction to the main ML techniques employed in the analysis presented in this Thesis.

5.2.1 Boosted decision trees

Decision Trees (DT) [128] are one of the most simple "traditional" machine learning algorithms. Their strength lies in their straightforward interpretability, as well as relatively fast training.

A decision tree consists of consecutive sets of binary selections (*nodes*) using observables (see Fig. 5.1). Each binary selection depends on formerly given answers and the final outcome (*leaf node*) of the tree is a classification of the given input, followed by a pre-defined maximum number of nodes.

Decision trees are first *trained* (i.e. optimising the selection) on samples of events with known labels, (x, y) , called *supervised learning*. As mentioned previously, many HEP problems consist of identifying signal processes, while rejecting background. The ML algorithm seeks to construct a function $F(x)$ which would approximate the "true" mapping of the data between x and y . In each node j , the working points c_j are varied while the efficiency of separation between signal and background are evaluated through the *loss function*, for instance, the *cross entropy function* [128] in the case of binary classifiers:

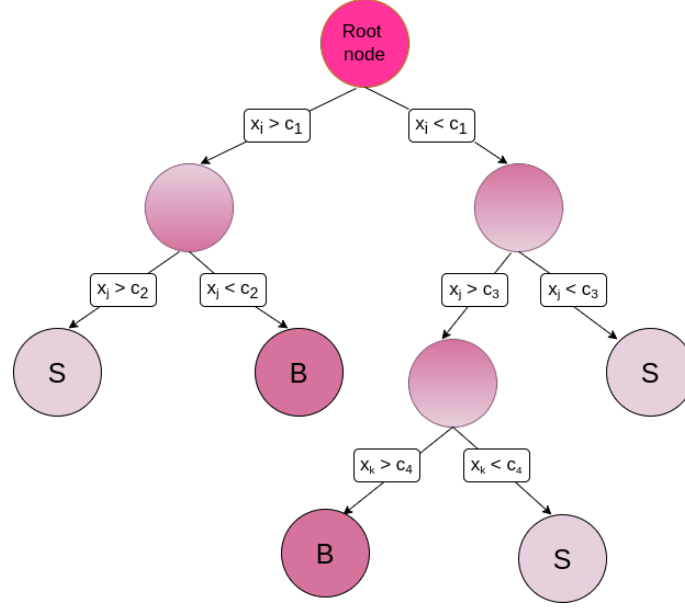


Figure 5.1: Schematic diagram of a simple binary decision tree where S is signal while B is background.

$$L = -\frac{1}{N} \sum_i^N (y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)) \quad (5.2)$$

where y_i is the binary indicator of the class label (0 or 1) of a single training example i , p_i is the predicted probability of the observation being in either of the classes, while N is the size of the training sample. To keep the nodes statistically significant, no further splits are produced below a pre-defined non-zero threshold of passing events.

However, single decision trees are not robust enough since they are easily subject to statistical fluctuations. Therefore, they can be grouped in ensemble methods, linearly combining multiple decision trees (*weak learners*) giving a pseudo-continuous final output based on the majority vote of all trees in the forest:

$$M(\vec{x}) = \sum_i^{N_{trees}} w_i \times h_i(\vec{x}; \vec{C}_i) \quad (5.3)$$

where each decision tree with binary classification output $h_i \in \{-1, 1\}$ is weighted by w_i , based on its performance. $\vec{x}$ are the set of data inputs in the ensemble method. While individual decision trees can have relatively low accuracy, the combined output of the majority vote will result in a *strong learner*.

There are various methods for training sequences of trees such as *boosting*, where the weights of misclassified events are increased, such that future learners enhance

their focus on them, while reducing their interest in correctly classified events. Such ensemble methods are called *Boosted Decision Trees* (BDT) [128, 129]. BDTs employ a gradient descent method to iteratively minimise the loss function with respect to the model parameters. Through the minimization process of the loss function, misclassified events are assigned larger gradients, hence the output of the *majority vote* would gradually tend towards y .

The major drawback of boosting methods is the possibility of *overtraining* or *overfitting* the random forests, such that the decision trees become sensitive to statistical fluctuations. This results in accurately learning the details of the given training set but leads to poor generalisation. *Pruning*¹ is an effective approach for keeping overtraining under control, as well as stopping BDT iterations at an optimal point, beyond which the algorithm would not improve in its performance but rather just overtrain itself (see Fig. 5.2). In addition, the *learning rate* of the BDT can be controlled by the *shrinkage* parameter to slow down the contribution of trees added in later iterations to the BDT.

During the training phase, half of the data sample is usually split into training and test samples (typically an 80%-20% split). The test sample serves to study the performance of the selection requirements learnt by the BDT during the training phase. The validation is usually quantified by the same loss function, as the one used for training, for which the training samples should not outperform the test samples. Additionally, once the BDT model is finalised on the training and testing data, it is used with a statistically independent validation data set. The training and testing data are discarded for future studies, such as statistical inference. The typical split on the overall available data set for training, testing and validation is 40%-10%-50%. The gradient boosting algorithm employed throughout this Thesis is called **XGBoost**, providing state-of-the-art performance [130].

The performance of a BDT classifier can be assessed by the *Receiver Operating Characteristic* (ROC) curve proving an efficient method for visualising the discrimination power of a given BDT model [128]. At each point on the ROC curve, the *true positive rate* (TPR) corresponding to the signal efficiency at a specific classifier output threshold, c , and the *false positive rate* (FPR) corresponding to the background rejection are evaluated:

¹cutting away insignificant nodes

$$\text{TPR} = \frac{S(x|c)}{S} \quad (5.4)$$

$$\text{FPR} = \frac{B(x|c)}{B}, \quad (5.5)$$

where $S(x|c)$ and $B(x|c)$ denote the number of signal and background events passing the classifier output threshold c . Another quantitative measure used during model optimisation is the purity associated with a specific selection. Purity is the fraction of correctly classified signal events divided by the sum of the signal and misclassified background (i.e. contamination) events: $p = N_{sig}^{pass} / (N_{sig}^{pass} + N_{bkg}^{pass})$.

Another method to quantitatively evaluate the performance of a BDT to other methods is the area-under-the-curve (AUC), defined as the area between the ROC curve and the x-axis, which represents the zero signal efficiency. Hence, it ranges between 0 and 100%, where the maximum value corresponds to perfect classification of signal and background events. A benchmark AUC value of 0.5 corresponds to a random classifier.

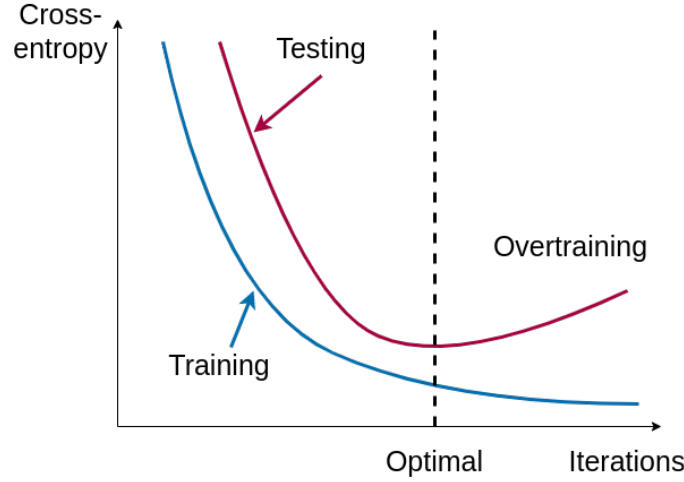


Figure 5.2: The cross-entropy validation function development the training and testing data samples as the number of BDT training iterations is increased. There is an optimal cut for the iteration number to maximise training on the data sample while minimising overtraining.

5.2.2 Neural Networks

Historically, artificial neural networks (NN) have been proposed to model biological neural systems [131]. Since the advancements of technology, NNs have been developed and employed in a wide range of applications, becoming cutting-edge technology.

The central element of an NN is an artificial neuron that calculates the *response* y to a given vector of *inputs* $\vec{x}$:

$$y = h\left(\sum_{i=1}^n w_i x_i + w_0\right) \quad (5.6)$$

where $\vec{w}$ are the *weights* imposed on each input element, w_0 is the *bias term* and h is the *activation function* of the neuron. Usually, the shape of the activation function is step function-like, for instance rectified linear unit (ReLU) $f(x) = \max(0, x)$, sigmoid $f(x) = 1/(1 + e^{-x})$ or hyperbolic tangent $f(x) = \tanh(x)(e^x - e^{-x})/(e^x + e^{-x})$, in order to learn non-linear correlations. The output layer uses a *softmax* function to transform the output vector into a vector of class probabilities ranging between 0 to 1:

$$\hat{y}_k = \frac{e^{z_k}}{\sum_j e^{z_j}}, \quad (5.7)$$

where z_k is the input to node k in the output layer and N is the total number of classes.

The most commonly used NN architecture is the *multilayer perceptron* (MLP) consisting of an *input layer*, *output layer* and single or multiple *hidden layers* [128, 132]. Each neuron in the NN receives input from all neurons in the previous layer and passes its output to all neurons in the next layer. The input to the first layer of neurons is typically the raw input data, whereas the output from the last layer of neurons is the network's prediction. The more hidden layers a network has, the deeper it is, and hence the more complex (see Fig. 5.3). MLPs can also be referred to as a fully connected, feedforward or *dense neural network* (dNN).

Another type of commonly used NNs are *convolutional neural networks* (CNN), which are specifically designed for processing and classifying data that has a grid-like structure, such as images, histograms or time-series data. The key feature of CNNs is their ability to learn features from the input data using convolutional layers, by applying filters (also called *kernels* or feature detectors) to small regions of the input data, and then pool or downsample the output to reduce its dimensionality. The study presented in Chapter 6 employs one-dimensional CNNs on a sequential series of tracks that have several kinematic and quality features associated with them.

Neural networks are typically constructed via supervised learning on a training data set, but not necessarily. Unsupervised learning algorithms also exist and are

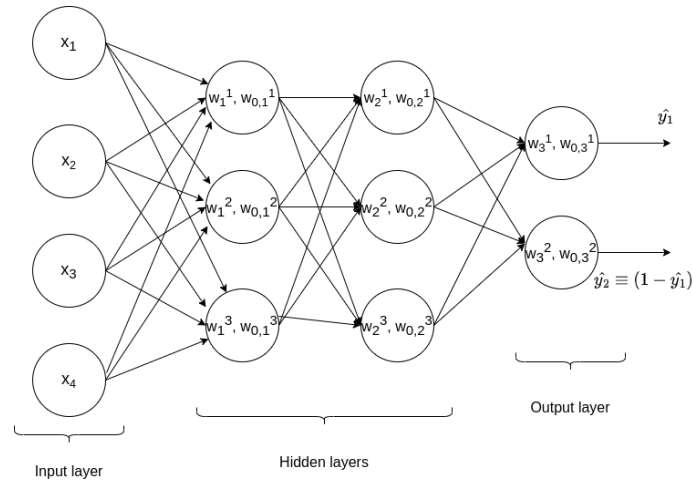


Figure 5.3: Schematic sketch of a multilayer perceptron or dense neural network with one input layer, two hidden layers and an output layer of two nodes corresponding to a binary classifier model.

used for clustering problems, for example *autoencoder algorithms* [128], *Gaussian Mixture Models* [128] or *Density-Based Spatial Clustering of Applications with Noise (DBSCAN)* [133]. During the training procedure, a loss function (typically a *mean square error* (MSE) function or cross-entropy) is minimised with respect to the network weights. The minimisation of a loss function is not trivial, therefore it is based on an iterative method using gradient descent, called (*error*) *backpropagation*. As a result, the NN model weights are updated in each layer. Unfortunately, the training of an NN can easily become exhaustive, hence overfitting may occur. To control the extent of overfitting, regularisation techniques are employed. As in the case of BDTs, connections between nodes in the NN may be randomly "dropped" during training, called *dropout*. To measure the performance of a NN during its training phase, a test set is provided over which the NN model is run. If the loss of the test set starts to increase after a number of training iterations, the NN model might be overtraining and is typically stopped at that point.

There is no clear answer whether NNs are better than BDTs or vice versa, the choice of algorithm should be made based on the type of problem to be solved. While BDTs are robust and easier to interpret, their main disadvantage is their inability to properly predict complex-shaped functions. BDTs need to be kept shallow to prevent overtraining, otherwise, the algorithm can become sensitive to outliers.

On the other hand, the advantage of NNs is their efficiency in predicting or approximating complicated functions. BDTs work well with a relatively small number of input features, whereas NNs are capable of processing hundreds or even thousands of input features, as the network structure can be extended both in width and

depth (called *deep learning*) in order to improve its performance, without the risk of overfitting. However, their results are rather cumbersome to interpret and involve finding the relevant correlations between input parameters.

5.2.3 Domain adaptation

Domain adaptation (DA) is an ML method employed in the DNN displaced jet tagger with the aim of addressing modelling differences between simulated input features and those from collected p-p data. As mentioned in Sec. 4.2.3, about 700 input features are fed into the DNN tagger, making it impossible to manually check the correct modelling of all of them via data to MC plots in all possible parameter spaces considered. However, there could be substantial differences between the simulated input features and data, given the limitations of MC generators, inaccurate detector modelling or difficulties modelling various particle interactions with the detector material. Typically, various correction scale factors are derived in different control regions to quantify the extent of mismodelling of physics variables, for example those presented in Chapter 9. However, such procedures would be unfeasible at the scale of input features used to train the DNN jet tagger.

Consequently, training the DNN displaced jet tagger to classify jets solely with mismodelled input features in the MC samples used would result in discriminator outputs that do not perform correctly in real p-p data. Potentially, correction scale factors could be derived a posteriori from the DNN tagger output to account for such mismodelling issues. Nevertheless, it is more desirable to train the DNN tagger a priori in a way that is insensitive to poor modelling in simulated samples.

The DNN displaced jet tagger employed in the analysis presented in this Thesis uses unsupervised DA by backpropagation [134] to account for modelling differences between simulated samples and collected data domains. The DNN tagger has two main branches as illustrated in Fig. 5.4: one branch is responsible for classifying the flavour of simulated displaced jets, called *class label predictor* and another branch for penalising the ML algorithm to learn the features of the simulated samples versus real data in the *domain label classifier* branch. The two branches are unified at the end of a feed-forward *feature extractor* branch, responsible for mapping the network input to a space of latent features that are fed into the two branches.

The domain label classifier branch is added to the NN architecture as a *gradient reversal layer* (GRL) and is trained via backpropagation. The additional loss term, L_d , quantifying the efficiency of the domain label classifier is added to the DNN, following the same form as Eq. 5.2, since the domain label classifier is a binary classifier

(simulation versus real data). The loss term of the multi-class label predictor is defined as:

$$L_c = -\frac{1}{N} \sum_i^N \sum_j^M y_i^j \cdot \log(p_i^j), \quad (5.8)$$

where j is the class label and M is the total number of classes in the class label predictor branch, i.e. the number of displaced jet flavour classes considered. The total loss function of the DNN tagger is the linear sum of the class label and domain label classifiers:

$$L = L_c + \lambda L_d, \quad (5.9)$$

where λ is a hyperparameter set to control the extent of DA and is optimised manually. Via this hyperparameter, the GRL ensures that the DNN penalises mismodelled input features while still enhancing the network's power to discriminate between jet flavour classes as a trade-off.

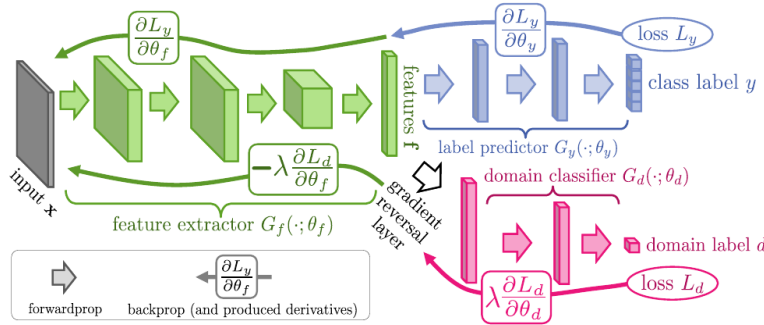


Figure 5.4: Schematic diagram of a NN employing domain adaptation by backpropagation via introducing a gradient reversal layer (bottom pink branch). The NN has a feed-forward architecture with a feature extractor (green) and class label predictor (blue) branch. The DA branch connects to the feature extractor branch and has the role of domain label classifier [134].

5.3 Statistical methods

This Section discusses the various statistical methods employed in the analysis presented in this Thesis. The first part of the Section discusses the quantification of correlation between two distributions, as well as uncertainty computations for the treatment of systematic uncertainties. Quantifying the extent of correlation between

physics variables becomes a relevant problem in the design of the event-level binary classifier presented in Chapter 8, as well as in the context of the ABCD background estimation method (see Sec. 5.3.6 and 10.1). The second half of the Section presents the framework for statistical inference employed to set upper limits on long-lived HNL model parameters.

5.3.1 The Pearson correlation coefficient

The Pearson correlation coefficient [135], ρ , is a quantitative variable that determines the extent of linear correlation between two variables x and y , with weights w :

$$\rho = \text{corr}(x, y; w) = \frac{\text{cov}(x, y; w)}{\sqrt{\text{cov}(x, x; w)\text{cov}(y, y; w)}}, \quad (5.10)$$

where $\text{cov}(x, y; w)$ is the covariance between variables x and y defined as:

$$\text{cov}(x, y; w) = \frac{\sum_i w_i (x_i - m(x; w))(y_i - m(y; w))}{\sum_i w_i}, \quad (5.11)$$

where $m(x; w)$ and $m(y; w)$ are the weighted means of variables x and y , respectively. Values of the Pearson correlation coefficient range between -1 to 1, negative values showing anticorrelation while positive values indicate a correlation between variables. The higher the absolute value of the correlation, the stronger the (anti)correlation between feature pairs.

5.3.2 The mutual information coefficient

The mutual Shannon information entropy or *M-score* [136], is a non-linear measure of statistical dependency between two random variables (observables), X and Y . In the discrete case, this is defined as:

$$M = I(X, Y) = H(X) + H(Y) - H(X, Y) = \sum_i \sum_j f_{X,Y}(x_i, y_j) \log_2 \left(\frac{f_{X,Y}(x_i, y_j)}{f_X(x_i)f_Y(y_j)} \right), \quad (5.12)$$

where $f(X, Y)$ is the two-dimensional joint probability density mass function and $f(X), f(Y)$ are marginal densities. The function $H(X) = -\sum_i f(x_i) \log_2 f(x_i)$

denotes the Shannon entropy of X . By definition, when the random variables are statistically independent, the joint density factorises $f(X, Y) \equiv f(X)f(Y)$ and the mutual entropy will be zero. This quantity therefore provides a stronger, more generic statement than linear correlation being zero.

Discretised densities of continuous observables are estimated via unit mass normalised histograms. The number of bins in two dimensions is chosen automatically using the *Scott rule* [137], and the histogram range per dimension is determined so as to contain 99% of the empirical probability mass to minimise the impact of outliers on binning. The maximum possible M -score value depends on the underlying distributions. To take this into account, a typical additive normalisation is used, which scales the maximum possible value of the M -score to unity:

$$I_N(X, Y) = 2 \frac{I(X, Y)}{H(X) + H(Y)}. \quad (5.13)$$

A simple exact relationship between the unnormalised mutual information, M , and Pearson linear coefficient, ρ , exists, assuming both marginal distributions are normally distributed:

$$M = -\frac{1}{2} \log(1 - \rho^2). \quad (5.14)$$

5.3.3 Statistical methods employed to estimate uncertainties

The following statistical uncertainty estimation methods are employed in the evaluation of the scale factors presented in Chapter 9, as well as in the calculations of statistical uncertainties of weighted MC events in histograms of kinematic observables and the ABCD background estimation methods used in the analysis discussed in this Thesis.

Clopper-Pearson binomial confidence interval

The Clopper-Pearson binomial confidence interval is a statistical method for estimating the range of plausible values for a binomial proportion, which is the probability of success in a sequence of independent trials with two possible outcomes [138]. The method is based on the idea of constructing a confidence interval for the binomial proportion using the binomial distribution:

$$\theta_{\text{low}} < \theta < \theta_{\text{up}} \equiv B\left(\frac{\alpha}{2}; k, n - k + 1\right) < \frac{k}{n} < B\left(1 - \frac{\alpha}{2}; k + 1, n - k\right) \quad (5.15)$$

where k is the number of passing events; n the total number of counts; α is the target error rate (i.e. the significance level); $\theta = k/n$ is the probability of success and $B(p, v, w)$ is the p^{th} quantile from the beta distribution with shape parameters v and w [139]. The Clopper-Pearson method provides a conservative estimate of the confidence interval by computing the lower and upper bounds of the interval using the inverse of the cumulative distribution function of the binomial distribution.

The lower and upper statistical uncertainties on θ are calculated as follows:

$$\sigma(\theta_{\text{low}}) = \theta - \theta_{\text{low}} = \frac{k}{n} - B\left(\frac{\alpha}{2}; k, n - k + 1\right) \quad (5.16)$$

$$\sigma(\theta_{\text{up}}) = \theta_{\text{up}} - \theta = B\left(1 - \frac{\alpha}{2}; k + 1, n - k\right) - \frac{k}{n} \quad (5.17)$$

At the limit case of 0 total counts ($n = 0$), the lower and upper uncertainties are set to 0. When the total counts are non-zero but 0 events pass ($k = 0$), the lower uncertainty is set to 0, while the upper uncertainty is set to $1 - (\frac{\alpha}{2})^{1/n}$. If $k = n$, the lower uncertainty becomes $\frac{k}{n} - (\frac{\alpha}{2})^{1/n}$ and the upper uncertainty is $1 - \frac{k}{n}$.

The statistical uncertainties for the scale factors derived in Chapter 9 are provided in App. A and are determined using the simple error propagation formula.

Uncertainty on weighted sum of events

Let $n = \sum_{i=0}^M w_i N_i$ be a weighted sum of event counts N_i with weight w_i of type i . The error on variable n is calculated as follows, assuming a Poisson error on the event counts $\sigma(N_i) = \sqrt{N_i}$:

$$(\sigma(n))^2 = \sum_{i=0}^M \left(\frac{\partial n}{\partial N_i} \sigma(N_i) \right)^2 = \sum_{i=0}^M (w_i \sqrt{N_i})^2 = \sum_{i=0}^M w_i^2 N_i \quad (5.18)$$

Pull values

The pull value is a measure of the correlation between two variables x and y with uncertainties σ_x and σ_y :

$$pull = \frac{x - y}{\sqrt{\sigma_x^2 + \sigma_y^2}} \quad (5.19)$$

The pull value indicates by how many standard deviations x deviates from y , taking into account the uncertainties on both measurements. Pull values close to zero (0-2) indicate that the two measurements show good statistical compatibility, while large pull values (typically above 3 or 4) may indicate a possible systematic effect or a measurement error, and can trigger a detailed investigation of the sources of the discrepancy.

5.3.4 Likelihood model

In HEP, the likelihood function is often used to determine the *parameters of interest* (POI) of a theoretical model that best describes the observed data. This is done by comparing the predicted distribution of events from the theoretical model with the observed distribution of events in the data. The expected number of events, $E[n_i]$, given the expected signal (s_i) and background (b_i) yields in channel i of a multi-channel experiment is defined as:

$$E[n_i] = b_i(\vec{\theta}) + \mu s_i(\vec{\theta}), \quad (5.20)$$

where μ is the *signal strength parameter*, the POI of the experiment and $\vec{\theta}$ is the set of *nuisance parameters* (NP) describing various systematic effects the signal and background yields depend on.

The likelihood function is defined as the probability of observing the data given the theoretical model and its parameters. Assuming that the number of experimental observations, n_i , are distributed according to a Poisson distribution [140, 141]:

$$\text{Poisson}(n_i | \lambda_i) \equiv e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!}, \quad (5.21)$$

where $\lambda_i \equiv E[n_i]$ is the mean expected value of the distribution, the Poisson likelihood model is constructed as follows:

$$\mathcal{L}_i(\mu, \vec{\theta} | n_i) = \text{Poisson}(n_i | b_i(\vec{\theta}) + \mu \cdot s_i(\vec{\theta})) \cdot p(\vec{\theta}), \quad (5.22)$$

where $p(\vec{\theta})$ is the prior distribution for the set of NPs, $\vec{\theta}$, assumed to follow a Gaussian distribution:

$$p(\theta_\kappa) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta_\kappa^2}, \quad (5.23)$$

where κ denotes a single NP from the set of NPs considered. Note that the likelihood function is not a probability density distribution of the signal strength parameter.

The goal of the likelihood model is to find the set of model parameters that maximise the likelihood function, called *maximum likelihood estimation* (MLE) [140, 141]:

$$L(\mu) = \max_{\vec{\theta}} \mathcal{L}(\mu, \vec{\theta} \mid \vec{n}), \quad (5.24)$$

that provides a conditional set of NPs that maximise the likelihood function $\mathcal{L}$, given the value of μ and all observed measurements, $\vec{n}$, in all experimental signal categories.

The likelihood model is particularly useful in HEP as it allows for the estimation of uncertainties in model parameters and predictions. This is achieved by constructing a likelihood function that includes the effects of statistical and systematic uncertainties in data and the theoretical model considered. The uncertainty on a given parameter is then determined by examining the change in the likelihood function when that parameter is varied within its uncertainty range.

5.3.5 Profile likelihood ratio

The *profile likelihood* is a statistical method that is more sensitive to the POI than the standard likelihood [140, 141]. In many cases, the POI is only a small part of a larger set of parameters that describe the system being studied. In such cases, it can be difficult to obtain precise estimates of the POI, as the other parameters can introduce significant uncertainty into the analysis. The profile likelihood addresses this issue by profiling over the other parameters, effectively integrating them out of the likelihood function. As a result, the dimensionality of the problem is reduced, allowing for more precise estimates of the POI. Furthermore, the profile likelihood has desirable mathematical properties that make it a well-behaved statistical method, allowing for accurate and reliable calculations of confidence intervals and p-values.

The *profile likelihood ratio* is a normalised version of the profile likelihood:

$$\lambda(\mu) = \frac{\max_{\vec{\theta}} \mathcal{L}(\mu, \vec{\theta} | n)}{\max_{\vec{\theta}, \vec{\mu}} \mathcal{L}(\mu, \vec{\theta} | \vec{n})} \equiv \frac{\mathcal{L}(\mu, \hat{\vec{\theta}}(\mu) | \vec{n})}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}} | \vec{n})}, \quad (5.25)$$

where $\hat{\vec{\theta}}(\mu)$ is the conditional maximum likelihood estimator of $\vec{\theta}$ in μ , i.e. for any value μ where $\vec{\theta}(\mu)$ is maximal, $\hat{\vec{\theta}}(\mu)$ is obtained. This process is repeated for all values of μ and divided by the overall maximum unconstrained likelihood, $\max_{\vec{\theta}, \vec{\mu}} \mathcal{L}(\mu, \vec{\theta} | \vec{n})$ to obtain the profile likelihood ratio that is only dependent on the POI, $\lambda(\mu)$. In Eq. 5.25, $\hat{\mu}$ is the maximum likelihood estimator of μ and $\hat{\vec{\theta}}$ is the unconstrained maximum likelihood estimator of $\vec{\theta}$.

The negative logarithm of the likelihood ratio is typically used and referred to as the *log-likelihood*, $\ell(\mu) = -\ln(\lambda(\mu))$ so that the function becomes a minimisation problem. According to the *Neyman-Pearson lemma* [142], the profile likelihood ratio test has the highest statistical power for a given test statistic, hence the highest signal purity.

5.3.6 The ABCD method

The ABCD background estimation method is a useful technique employed in HEP searches to estimate the background contribution in the (blinded) signal region in cases where the background is difficult to model or estimate directly from simulation [143]. The basic idea of the ABCD method is the choice of two statistically independent observables for the background processes (X, Y), that offer discrimination between the signal and background classes. The method is based on the observation that if two variables are statistically uncorrelated, then the probability density function of their product can be factorised into the product of their individual probability density functions:

$$\rho(x, y) = p(x)q(y), \quad (5.26)$$

in other words, the background distribution of events in one variable should be independent of the distribution of events in the other variable.

In the ABCD method, the data is divided into four regions based on two variables, A, B, C, and D. One of the regions (typically region D) is chosen as the (blinded) signal region. The remaining three regions (A, B, and C) are used as control regions to estimate the expected background contribution in the signal region:

$$\tilde{N}_D^{\text{bkg}} = N_B^{\text{bkg}} \times N_C^{\text{bkg}} / N_A^{\text{bkg}}, \quad (5.27)$$

where N_i is the number of observed events in regions $i = A, B, C$. The phase-space division of the ABCD background estimation method is schematically illustrated in Fig. 5.5.

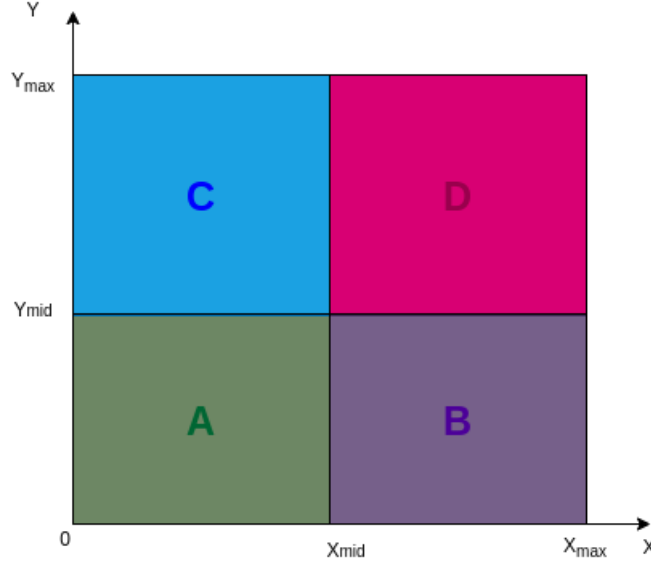


Figure 5.5: Schematic diagram of the ABCD background estimation technique definition where D is the blinded signal region of the analysis.

In practice, a likelihood model is constructed by separately considering terms for each SR category while incorporating the background prediction of the ABCD method, as well as the signal contamination in sidebands A, B and C in the likelihood fit:

$$\mathcal{L} = \prod_{i=48} \mathcal{L}_i^a(\vec{\theta}) \mathcal{L}_i^b(\vec{\theta}) \mathcal{L}_i^c(\vec{\theta}) \mathcal{L}_i^d(\vec{\theta}), \quad (5.28)$$

$$\mathcal{L}_i^m(n_i^m | \vec{\theta}) = \text{Poisson} \left(n_i^m \mid b_i^m(\vec{\theta}) + \mu \cdot s_i^m(\vec{\theta}) \right) \cdot p(\vec{\theta}), \quad m = a, b, c \quad (5.29)$$

$$\mathcal{L}_i^d(n_i^d | \vec{\theta}) = \text{Poisson} \left(n_i^d \mid \frac{b_i^b \cdot b_i^c}{b_i^a} + \mu \cdot s_i^d(\vec{\theta}) \right) \cdot p(\vec{\theta}), \quad (5.30)$$

where μ is the signal strength parameter, s_i^m and b_i^m are the expected signal and background yields in category i , region m , respectively. Assuming that the signal and background event counts in all four regions follow a Poisson distribution, the uncertainties in regions A, B and C are Poissonian: $\sigma_m = \sqrt{N_m}$ where $m = A, B, C$. Consequently, the pre-fit uncertainty on the expected event counts in region D is derived via the error propagation formula (Eq. A.1):

$$\tilde{\sigma}_D = \tilde{N}_D \sqrt{\frac{1}{N_A} + \frac{1}{N_B} + \frac{1}{N_C}}. \quad (5.31)$$

The ABCD method relies on the assumption that the two variables used to define the regions are uncorrelated, and it can be sensitive to systematic uncertainties related to this assumption. Thus, careful checks and validation of the method are required to ensure its reliability.

5.3.7 Hypothesis testing

A test statistic is a mathematical function that can be used in statistical hypothesis testing to evaluate whether the observed data is consistent with a given hypothesis or not. The test statistic is a numerical summary of the data that is calculated based on a statistical model and is used to compare the observed data with what would be expected under the null hypothesis (i.e., the hypothesis being tested). The choice of test statistic should be based on the properties of the data and the hypothesis being tested, as well as any assumptions or limitations of the statistical model.

In hypothesis testing, the null hypothesis (H_0) typically represents the status quo (SM, background-only) or the absence of any new effect, while the alternative hypothesis (H_1) represents the possibility of an alternative effect (signal-plus-background). The test statistic measures the degree of discrepancy between the observed data and the null hypothesis, and this discrepancy is used to calculate a *p-value*, which is the probability of observing a test statistic, x as extreme or more extreme than the observed value, assuming the null hypothesis is true:

$$p = \int_{\hat{x}}^{\infty} f(x) dx, \quad (5.32)$$

where $f(x)$ is the probability density function and $\hat{x}$ is the measured test statistic value. Often, the p-value is expressed in terms of the number of standard deviations of the normal distribution above which the integrated probability distribution corresponds to the value of a pre-specified threshold called *significance level*, p_0 [144]:

$$p_0 = 1 - \Phi(Z), \quad (5.33)$$

where $\Phi(x)$ is the cumulative distribution function of the normal distribution and Z is the required significance. For setting *upper limits* (ULs), the typical required

significance level is $p_0 = 0.05$, corresponding to the 95% *confidence level* (CL) and a significance of $Z \simeq 1.64$ standard deviations.

If the p-value is below the significance level, it suggests that the observed data are unlikely to have occurred by chance under the null hypothesis and that there is strong evidence against the null hypothesis. This can lead to the rejection of the null hypothesis in favour of an alternative hypothesis. Conversely, if the p-value is large, it suggests that the observed data are likely to have occurred by chance under the null hypothesis and that there is weak evidence against the null hypothesis. This does not necessarily mean that the null hypothesis is true, but rather that the observed data are consistent with the null hypothesis. It is important to note that the p-value is not a direct measure of the size or practical significance of the effect or association, or the probability of the null hypothesis being true or false, but rather a measure of the strength of evidence against the null hypothesis.

Upper limit setting

In HEP, upper limit setting is a statistical method used to set constraints on the existence or properties of New Physics, based on the observed data. Typically, for UL setting the definitions of the null and alternative hypotheses are reversed, such that H_0 becomes the background-plus-signal hypothesis whereas H_1 is defined as the background-only hypothesis with $\mu \leq \hat{\mu}$. The test statistic for UL setting is defined as [145]:

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \quad (5.34)$$

The UL on the model parameter can then be interpreted as a constraint on the existence or properties of the new particle or phenomenon, by comparing it with the predicted values from various theoretical models or scenarios. For example, if the upper limit on the cross-section or signal significance of a new particle is smaller than the theoretical prediction, it suggests that the particle may not exist or may have different properties than expected, and can guide the design of future experiments or the development of new theoretical frameworks.

Goodness-of-fit test

The *goodness-of-fit test* (GOF) is a statistical method that can be employed to evaluate the compatibility of a given model or probability distribution with the

observed data in the absence of an alternative hypothesis. In the case of the analysis presented in this thesis, a GOF test is used to test the compatibility between the observed data and the SM background-only hypothesis, i.e. the lack of HNLs observed. The saturated likelihood-ratio test statistic used in this test quantifies the level of agreement between the observed data and the null hypothesis [146, 147]:

$$q_{\text{sat}} = -2 \ln \lambda_{\text{sat}}(\mu, \vec{\theta}) = -2 \ln \frac{\mathcal{L}(\mu, \vec{\theta})}{\mathcal{L}_{\text{sat}}}, \quad (5.35)$$

where $\mathcal{L}_{\text{sat}}$ is the saturated likelihood constructed by artificially setting all NPs to their measured values. The saturated likelihood acts as a normalisation term, hence the resulting test statistic tends to a χ^2 distribution in the limit of large data. In the HNL analysis presented in this Thesis, the probability distribution of the statistic is generated using *MC toys*² and compared to the observed value in data. The large p-values observed between the toy MC distributions and the test statistic provide confidence in the compatibility of the observed data and the SM background-only hypothesis.

The asymptotic approximation

The *asymptotic approximation* [145] of the likelihood-ratio test statistic, $\tilde{q}_\mu$, relies on the asymptotic properties of the maximum likelihood estimates, which can be shown to be approximately normal under certain conditions, such as large sample sizes or when the number of parameters is small compared to the sample size. Specifically, the probability distribution of a likelihood-ratio test statistic $f(\tilde{q}_\mu | 0)$ can be shown to follow a χ^2 distribution if the null hypothesis is true, whereas for the case of an alternative hypothesis the probability distribution $f(\tilde{q}_\mu | \mu)$ follows a non-central χ^2 distribution.

Consequently, the asymptotic approximation of the likelihood-ratio test statistic allows for the computation of p-values and confidence intervals without the need for Monte Carlo simulations or numerical methods, which can be computationally expensive and time-consuming.

²artificially generated data using models that are often designed to mimic the basic characteristics of real data, but with known properties and without the complexities of the actual data

The Asimov data set

The Asimov data set is an artificially created "representative" data set that mimics the expected behaviour of the real data, without actually having to collect and analyse real data [145]. The Asimov data set is constructed by setting the values of the observed variables to their expected values under the specific hypothesis and adjusting the uncertainties to reflect the expected statistical fluctuations in the background. In the analysis presented in this Thesis, the simulated Asimov data set is employed in obtaining the expected ULs on various long-lived HNL models.

Discovery sensitivity

The *discovery sensitivity* is a quantitative measure used in the optimisation processes of event selection criteria and defining thresholds in the ABCD background estimation method of the analysis. The discovery sensitivity formula employed in this analysis is a result of the asymptotic approximation [148] applied on the Asimov data set [149]:

$$Z = \left[2 \left((s+b) \ln \left[\frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2} \quad (5.36)$$

where s, b are the expected signal and background yields, respectively, and σ_b is the uncertainty on the background prediction. Assuming $s \ll b$ and $\sigma_b \sim 0$, the expression reduces to $Z \sim s/\sqrt{b}$.

Nuisance parameter pulls and impacts

To test the effect of NPs on the best-fit value of the POI considered in the analysis presented in this Thesis, two statistical tools are used: *pull values* and the *impact* of the NPs on the best-fit value of the POI. By quantifying the effect of NPs on the POI, the robustness and reliability of the analysis can be assessed, ensuring that both statistical and systematic uncertainties are properly accounted for in the final results.

The NP pull values are defined similarly to Eq. 5.19:

$$\text{pull} = \frac{\hat{\theta} - \theta}{\sigma_\theta}, \quad (5.37)$$

where $\hat{\theta}$ is the best-fit value of the NP, θ is the initial (pre-fit) value, conventionally set to zero, and σ_θ is the input uncertainty.

The effect of the NPs on the POI influences the rate and shape of the POI. Varying the values of the NPs from their nominal values by $\pm 1\sigma$ deviated ones, their *impact value* is the difference in the POI estimate at $\pm 1\sigma$ deviation ($\hat{\mu}_{\theta \pm \sigma_\theta}$) and their mean estimate:

$$\text{impact} = \Delta\mu^\pm = \hat{\mu}_{\theta \pm \sigma_\theta} - \hat{\mu}. \quad (5.38)$$

The impact value can be interpreted as a correlation between the NPs and the POI: the higher the impact value of an individual NP, the stronger the effect on the best-fit value of the POI.

Chapter 6

Track-to-vertex association for the Phase-2 upgrade of the Level-1 trigger system

6.1 Introduction

As already introduced in Sec. 3.3, HL-LHC will increase the design luminosity of the LHC by a factor of between five and seven, which will enable scientists to perform more precise measurements of known particles, to search for NP, as well as to study rare processes. One of the main challenges faced by the CMS detector during the HL-LHC run will be to cope with the increased number of simultaneous collisions, as a result of the increased luminosity. Around 200 inelastic interactions are expected to occur in each p-p bunch crossing, resulting in increased PU conditions in the CMS detector. Consequently, the Level-1 trigger (L1T) system needs to be upgraded, as part of the Phase-2 upgrade of the CMS experiment.

The Phase-2 upgrade of the L1T will include information from the outer tracker modules for the first time, introducing track finding and primary vertex (PV) reconstruction in the L1T. Tracks with $p_T > 2$ GeV will be reconstructed by in-situ momentum measurement electronics, called *Level-1 time-multiplexed track-trigger* (TMTT) track finder algorithm [99, 150]. In order to efficiently reject pileup background, the aim of the L1T upgrade is to develop a fast and robust PV reconstruction algorithm that will correctly associate hard-scattering tracks with the PV, while separating them from PU. The goal of the work presented in this Chapter was the development of a binary classifier for track-to-vertex association in the L1T upgrade. The classifier will be employed downstream from a PV reconstruction

algorithm to determine whether a given track originated from the PV. The aim was to develop an *end-to-end neural network* for both tasks, as discussed in Sec. 6.4.

An additional challenge of the L1T upgrade is fitting the newly developed trigger algorithms, particularly the vertex reconstruction and association algorithm on custom hardware front-end FPGA electronics. A total latency of $12.5 \mu\text{s}$ will be allocated per bunch-crossing for L1T processing, of which about $5 \mu\text{s}$ will be allocated for the track-finding algorithm. Consequently, the design of the track-trigger algorithm must prioritise memory efficiency and latency.

The study discussed in this Chapter was produced in collaboration with the work presented in Ref. [151].

6.1.1 Baseline vertex reconstruction and track association algorithms

A proposed simplistic clustering algorithm for vertex reconstruction in the L1T is a grid-based binning method called *FastHisto* [76, 100, 152]. Reconstructed tracks are projected on the beamline, identifying the z-coordinate of the impact parameter, z_0 , they originate from. The z_0 position of tracks is filled in a 256-bin histogram ranging between -15 cm to +15 cm and weighted by their transverse momentum. Clusters are identified by looking at local maxima and finding a window around the position of the three consecutive bins with the highest combined p_T . The reconstructed PV corresponds to the middle of the three bins. FastHisto is a simple, quick, but rather crude method that can be implemented efficiently on FPGAs. One of its main issues is its failure to account for fake tracks with unphysically high p_T , accidentally associating them with the PV. In addition, it does not account for the worsening of the z_0 resolution at high η , leading to degraded PV resolution.

The baseline track-to-vertex association method is a *fixed window scan* (FW) around the reconstructed PV, applying a selection to the separation between each track and the vertex: $|\delta z| = |z_0^{\text{track}} - z_0^{\text{FastHistoPV}}|$, which is set as a function of the η to reflect the pseudorapidity dependence of the z_0 resolution. This method is reasonably effective, fast, simple, and easily implementable on the FPGA hardware. Nevertheless, it does not take into account complex track features and may incorrectly associate PU tracks in proximity to the reconstructed PV. Fig. 6.1 presents a visual illustration of the algorithm procedure.

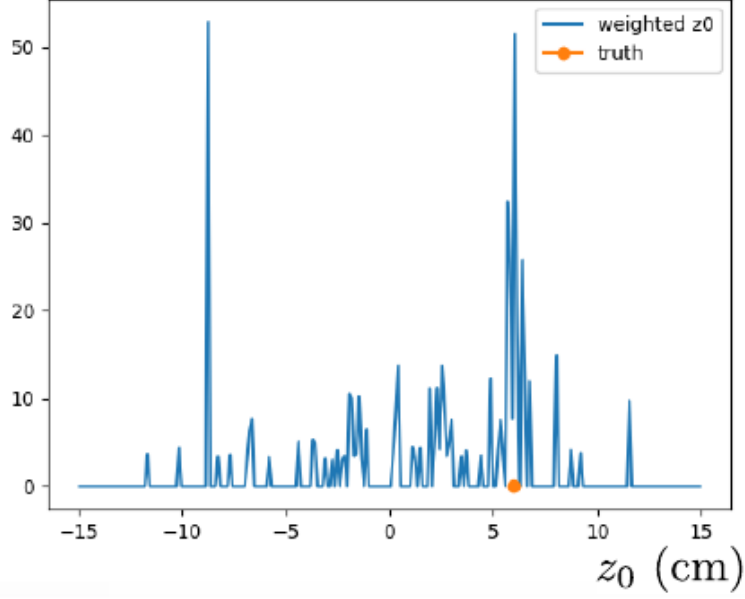


Figure 6.1: Event display of the FastHisto clustering algorithm, illustrating its vertex association in the z_0 position along the beamline. The true vertex in the event is contained within a fixed window since its associated tracks are reconstructed with a resolution. The left narrow peak is a "fake" track that was incorrectly constructed by the Kalman Filter by accidentally matching unrelated hits in the tracker. Its narrowness is a hint of its unphysical nature.

6.2 Training samples

The study presented in this Chapter was performed using two simulated data sets and employing a software simulation of the TMTT track-finder algorithm [99, 150] to reconstruct tracks from charged particles. Reconstructable tracks at the generator level are called *Tracking Particles* (TP), and the performance of the vertex finding and association algorithm is measured relative to tracks matched to TPs.

The simulation sets were generated at 14 TeV p-p collision energy with an average PU of 200 (PU200) interactions overlaying the hard-scattering process, describing the expected background environment at the HL-LHC. A maximum of 250 tracks per event were considered in the training of each track-to-vertex association algorithm described in Sec. 6.3. In order to reject fake tracks with unphysically large p_T , only tracks with $p_T < 500$ GeV were selected. The generator truth level information from tracks originating from the signal PV was flagged in the simulation set, and employed during the training of supervised learning algorithms, as well as the assessment of their performance.

A MC set of 100,000 $t\bar{t}$ events was employed for the training and validation of the track-to-vertex association algorithms investigated, as top pair production involves a

large invariant mass system with a decay resulting in a high multiplicity of high- p_T tracks. In addition, a set of 33,000 simulated $Z \rightarrow \mu\mu$ events were employed in the study due to the "clean" topology of the decay system with a well-defined PV and high- p_T tracks. Both systems have desirable properties for PV reconstruction and track-to-vertex association algorithm development efforts.

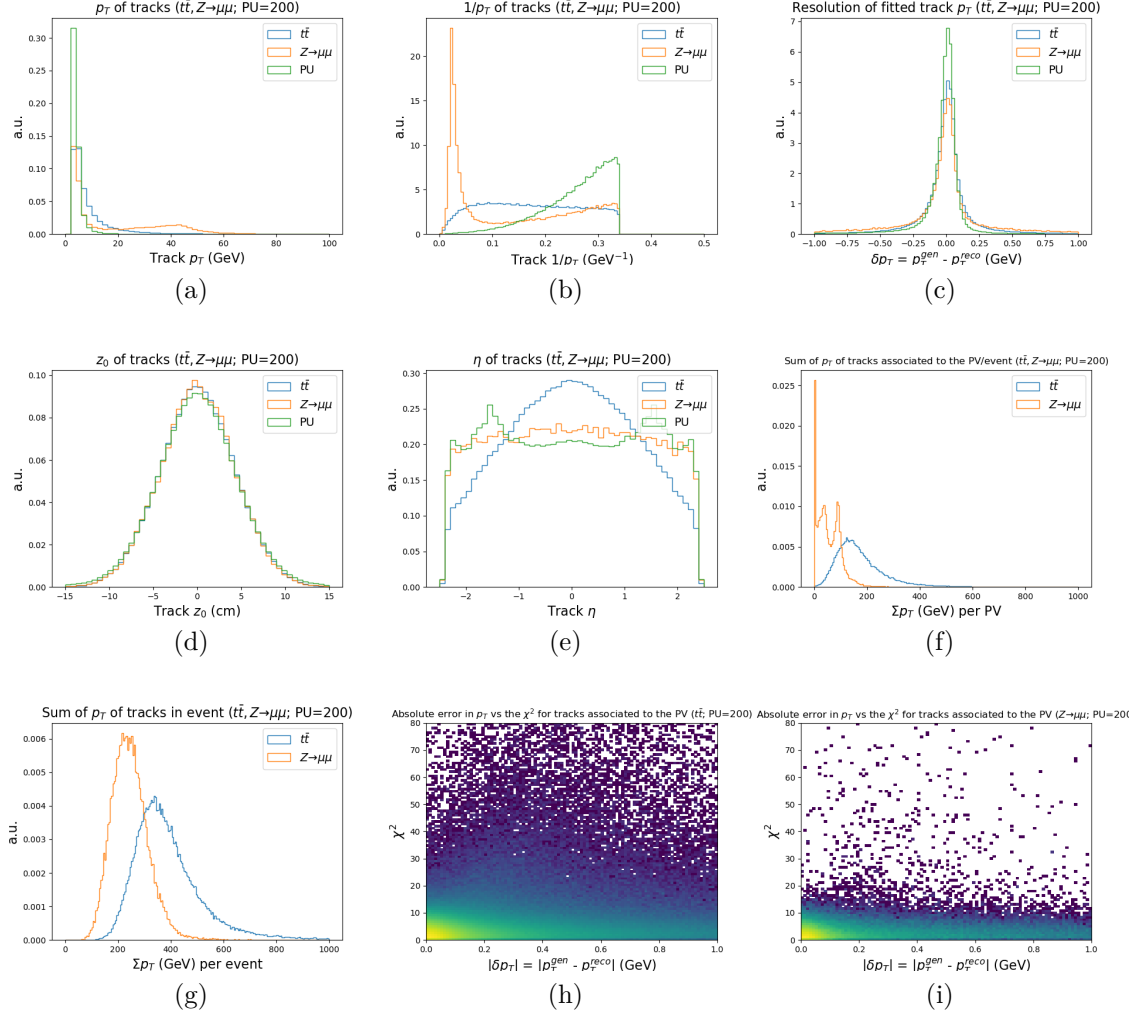


Figure 6.2: Track and event feature distributions of selected signal and PU tracks.

In order to train the binary classifiers for the track-to-vertex association, it is crucial to understand how various track features will perform as potential inputs, as well as understand their potential correlation. A number of possible input features describing the kinematics, position, orientation and fit accuracy of tracks were investigated for both signal and background tracks: z_0 , δz , p_T , η , $\tan \lambda$, χ^2 , p_T^2 and $1/R^2$, where χ^2 is the goodness of fit parameter given by the Kalman Filter for reconstructed tracks; $\tan \lambda \equiv \sinh \eta$ is a representation of the angle measured by the pseudorapidity including information on the azimuthal angle, since tracks should be uniformly distributed in ϕ ; $1/R^2$ is the curvature of the track, containing the same information

as p_T . Initial investigations showed that p_T^2 , $1/R^2$ and $\tan \lambda$ were redundant variables, because the performance of the BDT classifier described in Sec. 6.3.1 did not change when they were included as input features. Hence, they were disregarded in all subsequent analyses, since they contained overlapping information with p_T and η . Eventually, the inverse of the p_T was employed in training the binary classifiers due to memory efficiency considerations in the FPGA chips and being provided as an output of the Kalman filter. The remaining 5 input features were used to build the various binary classifier models described in Sec. 6.3.

Fig. 6.2 presents distributions of input features characterising the kinematics of both signal and background tracks, as well as overall event observables such as the summed p_T of tracks in an event and those associated with the PV. In addition, two heatmaps are presented showing the correlation between the p_T resolution and χ^2 of signal tracks.

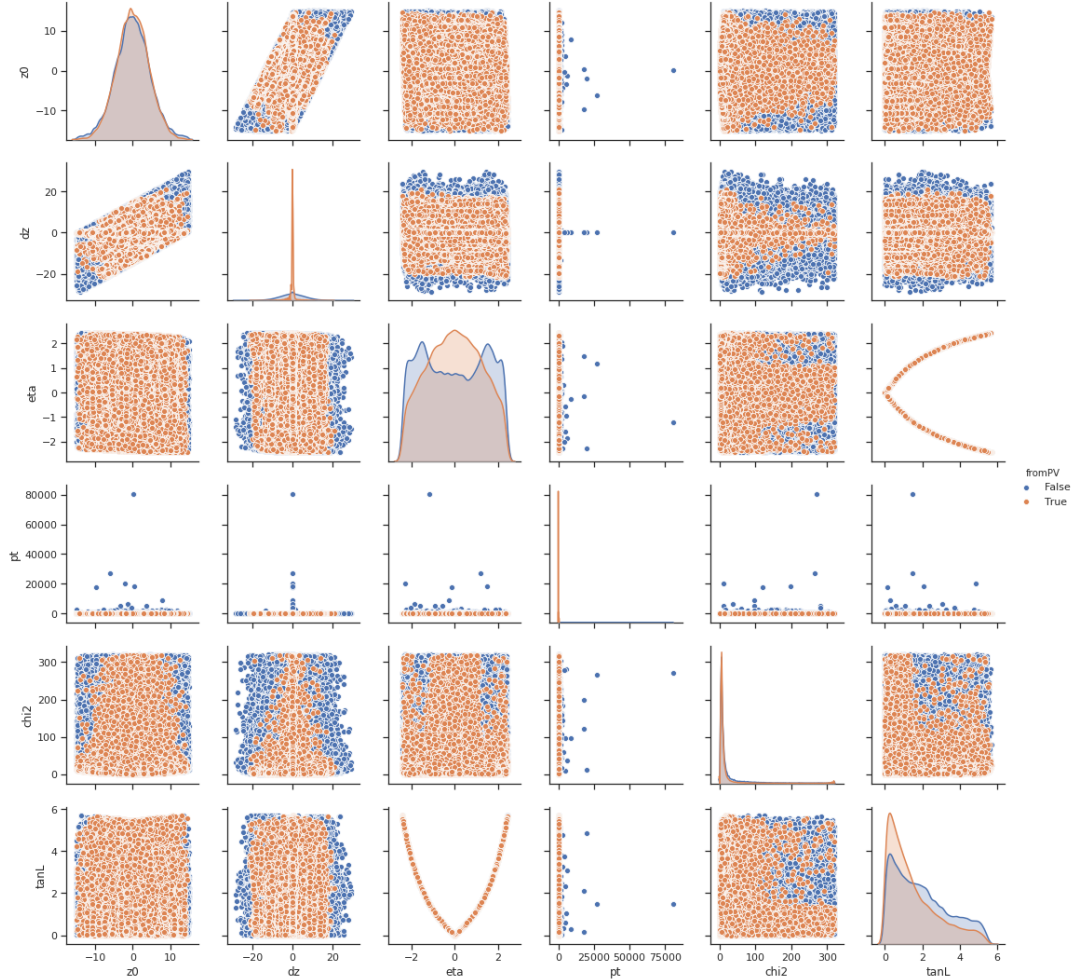


Figure 6.3: Quantifying feature correlations via scatter plots for $t\bar{t}$ ("true") and PU tracks ("false").

Possible correlations between 6 input features were quantified using scatter plots

and heatmaps to further understand the characteristics of the inputs used with the MVA methods considered. Fig. 6.3 presents the 2D distributions for $t\bar{t}$ and PU tracks for each feature pair, without showing their density in each 2D bin. There is a clear correlation between η and $\tan \lambda$, but otherwise, other distributions are rather symmetric demonstrating no correlation between various other features. Hard-scattering tracks from the $t\bar{t}$ events tend to occur at lower η values given the high invariant mass of the system, whereas the pseudorapidity is rather flat for the Z events. Pile-up tracks, on the other hand, peak at $|\eta|$ values of around 1.5. Furthermore, some PU tracks have unphysically high p_T values, since they are "fake" tracks constructed by accidentally associating unrelated hits in the tracker by the Kalman Filter. In addition, it can be observed that these tracks tend to cluster around $\delta z = 0$, while physical tracks have a uniform distribution in δz . Some of the associated heatmaps can be found in App. B, further proving no correlation between various feature pairs. The total p_T sum per signal vertex has a three-peak structure for $Z \rightarrow \mu\mu$ events reflecting whether both muons are reconstructed in the event, the $\sum p_T$ peaking at around the Z mass for such events. If no muon tracks are reconstructed, the $\sum p_T$ of the PV peaks at 0 GeV, whereas if one muon track is reconstructed, the distribution peaks at around 40 GeV.

6.3 Track-to-vertex association with MVA methods

6.3.1 Association using boosted decision trees

As discussed in the previous Section, a range of possible input variables describing track kinematics and position were considered for training the BDT model designed to separate signal tracks originating from the hard-scattering PV from tracks resulting from PU interactions. Initial studies considered the following set of six input features: z_0 , δz , p_T , η , $\tan \lambda$, χ^2 , while training several BDT models with varying number of input feature sets. The BDT algorithm of choice was `GradientBoostingClassifier` from the `sklearn` package [153] which was trained exclusively on the $t\bar{t}$ PU200 sample. For the initial investigations, only the default BDT parameters were used, as defined in the `sklearn` documentation: learning rate 0.1; number of estimators 100; minimum number of samples per leaf 1 and maximum tree depth of 3. A hyperparameter scan on the BDT parameters showed little dependence on the training parameters when evaluating the performance of the BDT model given a fixed set of input features. The loss function employed for the training was binary cross-entropy, the available

$t\bar{t}$ simulation sample being split 90%-10% between the training and testing phases.

The various BDT models were compared by their performance via their AUC and validation cross-entropy loss parameter. The extent of the correlation between input features could also be quantified by its impact on classification performance when added to the BDT model. It was found, as expected, that η and $\tan \lambda$ were strongly correlated, therefore the latter was dropped from the input feature set, aiming to keep the BDT model as simple as possible, given the latency limitations for its possible hardware implementation. In addition, for memory efficiency reasons (given no bias was observed in the input variable distributions), the absolute values of η and δz were used in training the final BDT model.

The performance of the BDT models with varying numbers of input features was compared using plots of the AUC versus the logarithmic loss (see Fig. 6.4). It was found that including δz and p_T in the BDT model was the most crucial in defining the performance of a model, rather than the total number of input features. Moreover, ordering the input features by their contribution to the model's decision-making power confirmed that, indeed, the aforementioned variables were the most discriminating ones, as shown in Fig. 6.5. The order difference in importance between input variables for the two signal samples considered is due to differences in their event topologies.

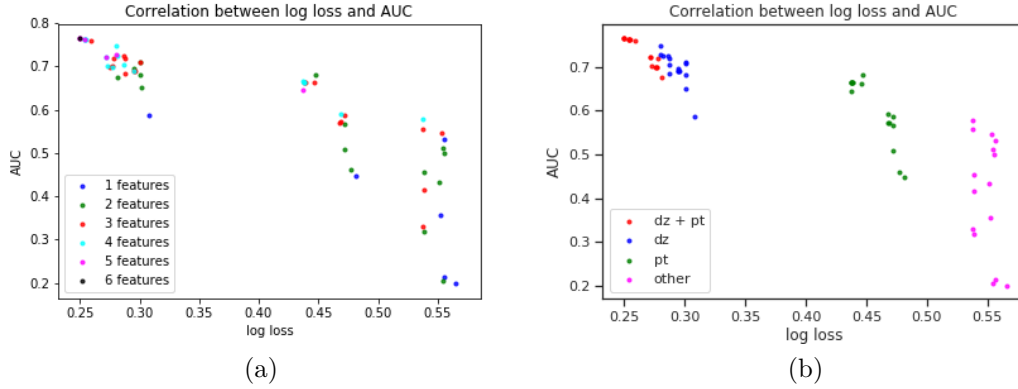


Figure 6.4: Graphs presenting the possible correlations between the AUC and the logarithmic loss for the various BDT models trained on the $t\bar{t}$ sample. In the left plot, the number of input features for individual models is illustrated in a different colour while the right plot shows the impact on the BDT performance of the inclusion of specific input variables, such as δz and p_T .

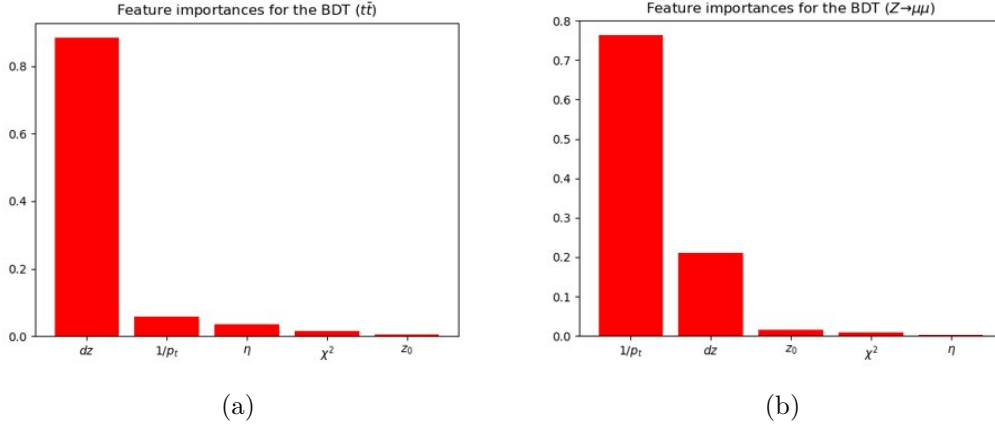


Figure 6.5: Quantifying relative feature importances in the BDT model for $t\bar{t}$ (left) and $Z \rightarrow \mu\mu$ (right) signal processes.

6.3.2 Association using dense and convolutional neural networks

Two neural network (NN) models were investigated as candidate algorithms for track-to-vertex association that could eventually be integrated into the end-to-end NN. Due to latency and memory limitations, the two networks needed to be designed with a relatively low number of parameters while maintaining reasonable discriminating performance.

Both a dense and a one-dimensional convolutional NN architecture were considered for the task (see Sec. 5.2.2). Each network had two hidden layers, each with 32 nodes and a ReLu activation function. A dropout of 10% was applied after each hidden layer in order to keep overtraining under control. The activation function of the last layer was a sigmoid, while the loss function optimised throughout the training was binary cross-entropy. A kernel size of 1 was applied in the CNN when training on an entire event, which was a requirement for the end-to-end NN implementation. A batch size of 400 events was trained for 20 epochs. The split between training and testing data sets was 90%-10%. The two networks were trained separately on tracks originating from $t\bar{t}$ and $Z \rightarrow \mu\mu$ events. The same input features were employed as for the BDT classification model: z_0 , δz , $1/p_T$, η , χ^2 . Figs. 6.6 and 6.7 show the progress of optimisation of the classification accuracy and the cross-entropy loss as the training of the dense NN (dNN) and CNN progressed by epoch. There is no evidence of overtraining in the loss curves and good separation between signal and PU tracks was observed.

A hyperparameter scan was performed to tune the network parameters for better

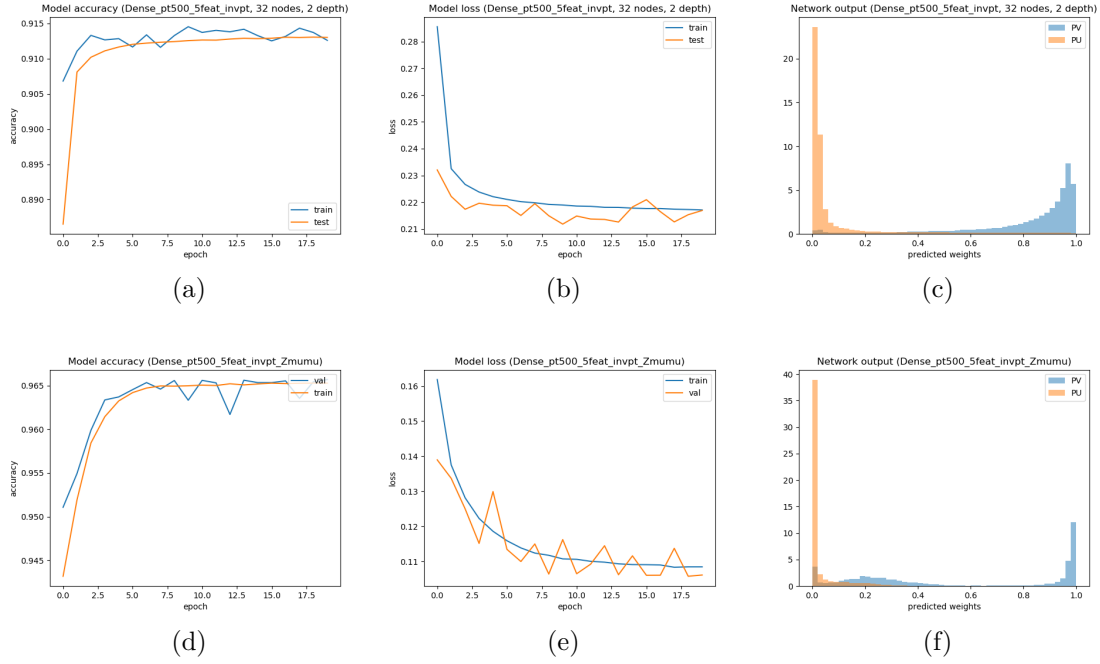


Figure 6.6: Classification accuracy (left) and cross-entropy loss (middle) development by training epoch of the dense NN model, as well as in the test data set. The dense NN output (right) shows good separation in the test set. Top plots show the dNN training and test performance for $t\bar{t}$ events, whereas the bottom was trained on $Z \rightarrow \mu\mu$ events.

performance, as well as different dNN and CNN network architectures were tested. Larger NN architectures showed only 0.5% improvement in performance, whereas latency and memory usage increased significantly.

6.3.3 Performance of the MVA models considered for track-to-vertex association

Two measures are used to evaluate the performance of the binary classifiers employed for track-to-vertex association. ROC curves are the traditional way to illustrate the performance of binary classifiers as discussed in Sec. 5.2.1. In addition, a *precision-recall curve* is typically used for L1T problems, reflecting the physical nature of the inputs and measuring the performance of the trigger itself. The physical requirements of trigger algorithms prioritise both the efficiency of correctly associating tracks to the PV, as well as the purity of the tracks being flagged as originating from the signal vertex while minimising mistagging. Therefore, purity vs matching efficiency curves were introduced to quantify the performance of the binary classifiers. *Purity* is defined as the fraction of reconstructed tracks associated with a reconstructed PV that actually matches the corresponding generator level PV: purity

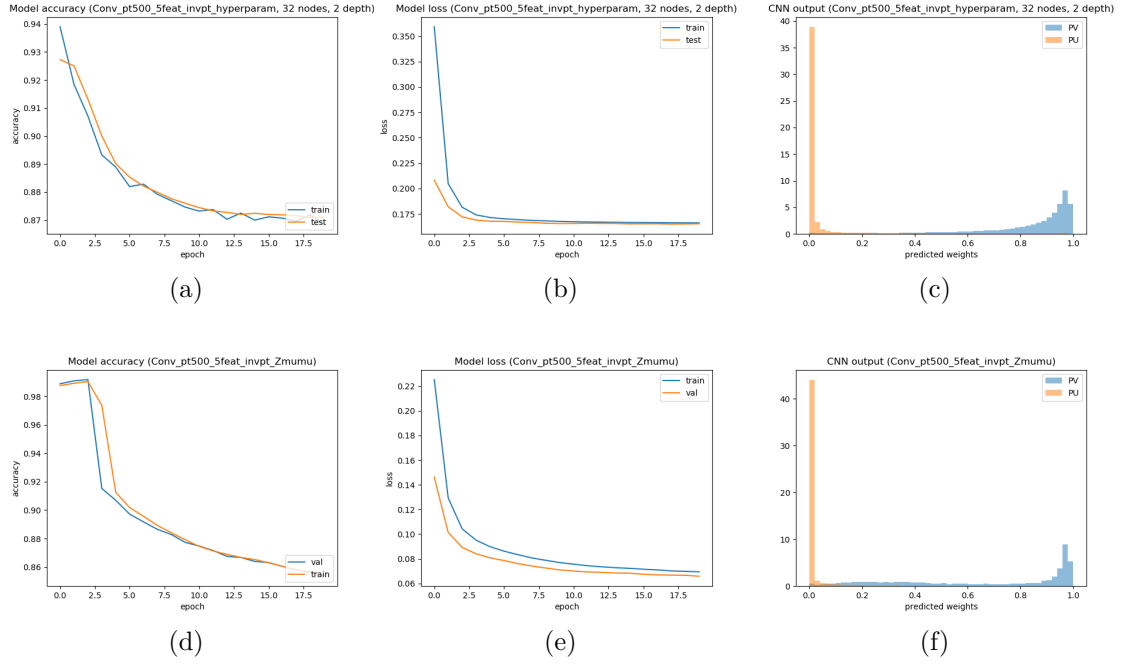


Figure 6.7: Classification accuracy (left) and cross-entropy loss (middle) development by training epoch of the convoluted NN model, as well as in the test data set. The CNN output (right) shows good separation in the test set. The top plots show the CNN training and test performance for $t\bar{t}$ events, whereas the bottom was trained on $Z \rightarrow \mu\mu$ events.

$= 1 - \text{contamination}$. *Efficiency* is defined as the fraction of reconstructable generator level tracks associated with a PV found in the reconstructed vertex: $\text{efficiency} = 1 - \text{missed tracks}$. In this study, the aim is to minimise the AUC from the ROC curves, while maximising it for the purity vs matching efficiency curves. Two more quantitative measures were also considered in evaluating the performance of the track-to-vertex association models: the *Characteristic True Positive* rate (R_C^{TP}) at which the *False Positive* (FP) rate is 1% and the *Characteristic Purity* value (P_C) at which the matching efficiency is 95%.

Fig. 6.8 shows the performance of the four track-to-vertex association models investigated in this study on both $t\bar{t}$ and $Z \rightarrow \mu\mu$ events. The fixed window scan systematically underperforms compared to all MVA methods in both signal topologies. The BDT and dNN show comparable performance, whereas the CNN model has the best performance from all track-to-vertex association models considered, being able to reject 99% of PU tracks while correctly associating 54% of $t\bar{t}$ and 69% of $Z \rightarrow \mu\mu$ tracks. Moreover, it provides 81% purity for $t\bar{t}$ tracks and 48% for $Z \rightarrow \mu\mu$ tracks at 95% efficiency. Nevertheless, the CNN model had 1441 parameters, adding much latency and memory usage for eventual implementation in FPGA hardware.

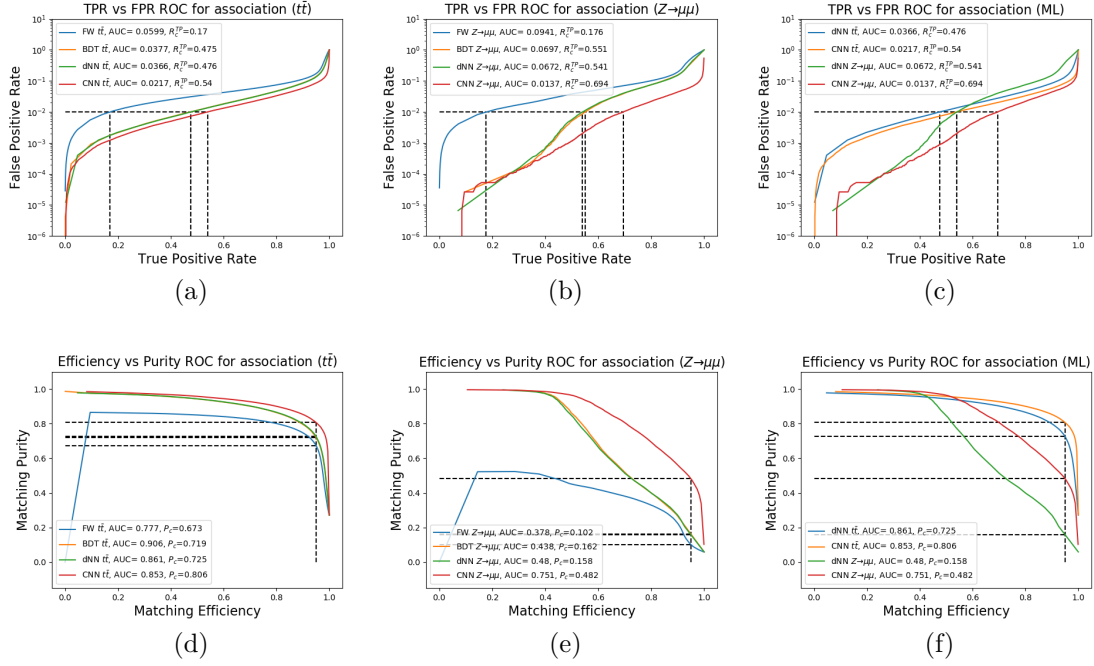


Figure 6.8: ROC curves (top) and purity vs matching efficiency (bottom) curves comparing the performance of the four track-to-vertex association models investigated in this study. The left plots show the training performance on the $t\bar{t}$ data set, whereas the middle ones on $Z \rightarrow \mu\mu$ events. The right plots compare the performance of the two NNs on both data sets.

6.4 End-to-end neural network for vertexing

In the context of the L1T upgrade, FastHisto served as the baseline algorithm while a proposed end-to-end neural network was developed to optimally identify the z_0 position of PVs and associate signal tracks with them [99]. The end-to-end nature of the NN consists of having several layers of separate NN sub-systems with different operational tasks, all connected together from the primary input data set to the final output. The advantage of the end-to-end NN architecture is the simultaneous optimisation of each processing step, instead of manually optimising them in isolation. However, end-to-end NNs yield larger and more complex architectures with additional challenges to validate their performance. Nevertheless, intermediate representations between processing steps can reveal significant correlations between the model inputs that may otherwise remain elusive. Thus, end-to-end models can serve as valuable aids in comprehending unfamiliar problem domains and identifying effective data representations for a particular problem.

The L1T upgrade aims to integrate the novel end-to-end NN into FPGA chips while designing it to perform better and more robustly than the previously proposed

FastHisto method. In addition, a design priority was keeping its latency and simplicity relatively low compared to other well-established clustering algorithms. The proposed end-to-end NN architecture is illustrated in Fig. 6.9. Histograms were created via 1D CNN using weighted event tracks, followed by a peak finding algorithm employing another layer of CNNs. The peaks in the weighted track histogram were interpreted as an estimation of the PV position. Besides identifying the PV position, the NN was trained to supply latent variables for inferring the track-to-vertex association probability, $P(\text{track}|z_0^{\text{PV}})$. A modified version of one of the proposed track-to-vertex association algorithms would then have been integrated at the final "convolutions" stage of the end-to-end NN.

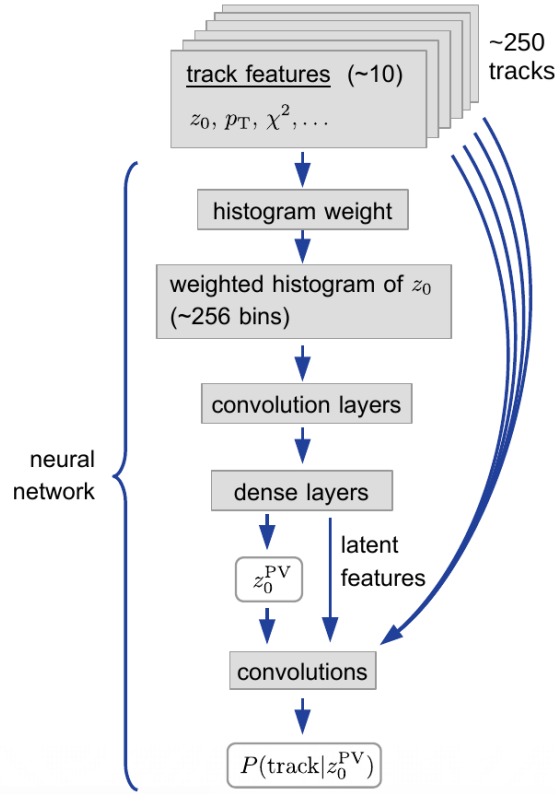


Figure 6.9: Schematic of the proposed end-to-end NN architecture for PV reconstruction and track-to-vertex association for the Phase-2 L1T upgrade [99].

6.4.1 Epilogue

Further studies were proposed for the continuation of the work presented above, particularly for the track-to-vertex association algorithms. Investigating the performance of the various algorithms introduced in Sec. 6.3 for vector-boson fusion (VBF) Higgs boson decays to invisible final states would have added a different topology of interest for the HL-LHC physics program. In addition, the final track-to-vertex

association models should have been trained on mixed signal samples, as only one algorithm could have been implemented in the end-to-end NN.

The track-to-vertex association models presented above were eventually integrated into the CMS offline software in order to study the improvement of the trigger efficiency compared to the case when FastHisto was used. The output of the CMS offline software showed some bias in the reconstructed z_0 of the PV. After careful studies, it was found that the issue resulted from a reconstruction bias in the TMTT track-finding algorithm used for the tracks that the algorithms were trained on. A new set of simulated tracks was produced via a hybrid track-finding algorithm.

Ultimately, the architecture of the end-to-end NN was slightly modified, including the inputs of a BDT algorithm that discriminates real tracks from fake ones. The z_0 position of the reconstructed PV was passed to the track-to-vertex association model consisting of a three-layer dense NN. Information from the BDT was also fed to the track-to-vertex association algorithm, along with the p_T and η of the reconstructed tracks. The end-to-end NN was successfully implemented in FPGA firmware and showed promising performance results compared to the cut-based approach considered previously [154].

6.5 Summary

This Chapter presented a study of track-to-vertex association algorithms developed as part of the Phase-2 upgrade efforts of the CMS Level-1 trigger. An end-to-end neural network was proposed to reconstruct the hard-scattering PV and to infer its position using deep learning techniques. In addition, the end-to-end neural network would associate signal tracks with the reconstructed PV. The design of the end-to-end neural network prioritised low latency and memory efficiency while retaining robust performance as it was implemented in FPGA firmware employed in the upgraded L1 trigger.

The proposed track-to-vertex association algorithms showed significant performance improvement compared to the cut-based approach originally proposed. The end-to-end NN was successfully integrated into FPGA hardware, showing promising results and improvement in performance in comparison to the baseline model.

Chapter 7

Search for long-lived heavy neutral leptons with the CMS detector

This Chapter presents the analysis strategy employed to search for long-lived HNLs in data collected by the CMS experiment between 2016-2018. Sections 7.1 and 7.2 present the simulated signal models and relevant SM background processes used in the analysis. Section 7.3 discusses the trigger strategy used to collect possible signal events. Section 7.4 introduces the various physics objects and their selection criteria. Finally, Sec. 7.5 provides a brief introduction to a novel deep neural network-based (DNN) displaced jet tagger (Ref. [9]) that aims to identify the displaced topology of HNL signal decay products.

In addition to the displaced jet tagger, the analysis employs an event-based classifier that aims to capture the higher-level kinematics of signal events, which forms the subject of Chapter 8. The sources of systematic uncertainty considered in this analysis are discussed in Chapter 9. The results of this analysis are then presented in Chapter 10, employing a data-driven background estimation technique using the previously mentioned event- and jet-level discriminants.

The analysis strategy presented in this Chapter is based on the work of an analysis team that the author of this Thesis was part of and made contributions to. Consequently, certain figures and ideas in this Chapter are based on works previously published in Ref. [109].

7.1 Heavy neutral lepton signatures

Long-lived HNLs are an attractive and well-motivated BSM scenario to investigate in experiments at the LHC, as discussed in Secs. 2.2.2 and 2.3.2. There is still a large part of the mass-coupling parameter space uncovered by previous searches, which this analysis has the potential to probe. In this analysis HNL masses in the range 1-24 GeV¹ with lifetimes ranging between 10⁻⁵ mm up to 10 metres are sought. Both Dirac and Majorana-type HNLs are considered, both of which can induce *lepton flavour violation* (LFV), whereas the latter can produce *lepton number violation* (LNV), as well. If HNLs were discovered, the result could shed light on whether HNLs are Dirac-type or Majorana-type particles. This analysis considers only one active HNL with couplings V_e , V_μ , and V_τ to the three lepton generations, respectively.

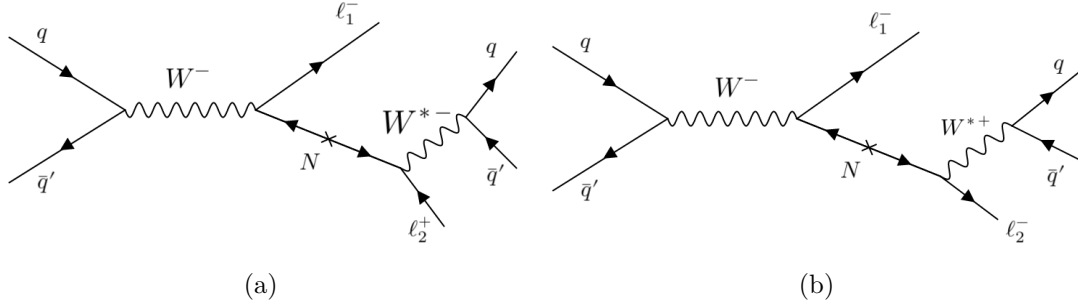


Figure 7.1: Feynman diagrams presenting the CC HNL production and semi-leptonic decay channels considered in this analysis. Charge-conjugate modes of the on- and off-shell bosons are also considered. The left diagram shows an opposite-sign dilepton final state available for both Dirac and Majorana HNL scenarios, whereas the right diagram shows a same-sign dilepton final state possible only for Majorana-type HNLs. The latter is an instance of LNV.

Searching for HNL signatures with the CMS detector has its own challenges, particularly for long-lived scenarios. Firstly, long-lived HNLs typically produce soft, low- p_T decay products that are difficult to trigger on. Therefore, to capture HNL signal events, additional particles also need to be produced that can be used for triggering. While there are a number of potential HNL production scenarios possible in CMS (listed in Sec. 2.2.2), this analysis focuses on the *charged-current* (CC) production mode, which results in a prompt, energetic lepton (ℓ_1) that the event is triggered on, and a neutrino that mixes with the HNL. The HNL then decays off-shell via a W boson into a displaced, soft secondary lepton (ℓ_2) and a pair of boosted quarks ($q\bar{q}'$) that collimate into a displaced jet, captured by the DNN displaced jet tagger discussed in Sec. 7.5. Two HNL production and decay modes are shown in Fig. 7.1,

¹Though derived limits are produced for HNL masses up to 20 GeV.

indicating both *opposite-sign* (OS) dilepton final states produced by both Dirac and Majorana-type HNLs, as well as *same-sign* (SS) dilepton final states produced only by Majorana-type HNLs, being an instance of LNV. Both processes can produce dilepton states of differing flavours, being a case of LFV. In the case that the HNL is a Majorana particle, its partial width would be twice as large as the case for Dirac HNL particles:

$$\Gamma_{\text{Majorana}} = \Gamma_{\text{Dirac}}^N + \Gamma_{\text{Dirac}}^{\bar{N}} = 2 \Gamma_{\text{Dirac}}. \quad (7.1)$$

HNL production via B and D meson decays is not considered in this analysis, since the decay products associated with the HNL are soft tracks that are challenging to trigger on in CMS. Moreover, such decays are limited by the mass of the parent meson, hence limiting the potential HNL masses probed (typically below 1 GeV). Since such HNL particles are soft, they decay via two-body decays, resulting in a displaced lepton and a quark pair that eventually forms a light meson: $N \rightarrow \ell_2 X$, where X denotes a light meson. A dedicated search is needed to investigate such scenarios.

Furthermore, HNL production via Z or Higgs boson decay is discarded in this search since such processes would need to trigger on the associated SM neutrino produced along with the HNL particle. Neutrinos do not interact with the detector material, therefore they manifest as large missing energy in collision events. However, it is very challenging to trigger on small amounts of missing energy in CMS events.

This analysis focuses exclusively on semi-leptonic decays of HNLs aiming to fully reconstruct the HNL with visible physics objects in the detector. Semi-leptonic HNL decays have branching ratios almost twice as high as fully leptonic ones. Despite the latter are considered to have "cleaner" event signatures, the DNN displaced jet tagger has the potential to accurately identify both the hadronic part and displaced lepton (if merged into the jet) of the HNL decay.

In conclusion, the HNL production and decay channels searched for in this analysis proceed exclusively via CC, providing a semi-leptonic final state via three-body decay of the HNL:

$$pp \rightarrow W \rightarrow \ell_1 N \rightarrow \ell_1 \ell_2 W^* \rightarrow \ell_1 \underbrace{\ell_2 q \bar{q}'}_{\text{displaced jet}} \quad (7.2)$$

This analysis considers multi-generational mixing, coupling the HNLs to all three generations of SM neutrinos, as opposed to the single-generation couplings previously

considered in such searches. In fact, based on the ν MSM and Fig. 2.5 in Sec. 2.2.2, simultaneous couplings to multiple SM neutrino generations are more likely for both Dirac and Majorana HNL scenarios. The HNL coupling to tau-neutrinos is still relatively unexplored (Fig. 2.2c in Sec. 2.2.2), posing an interesting opportunity for a break-through application of the DNN jet-tagger. Nevertheless, this channel has its own challenges, since if tau leptons decay leptonically, they always emit a neutrino, which results in the detectable daughter lepton being soft because the associated neutrino carries away most of the tau lepton's energy. Therefore, triggering on events with the leading lepton being a tau is difficult and inefficient. Hadronic tau decays could be investigated, however, the lowest single-tau trigger thresholds in CMS surpass the kinematic limits for tau particles resulting from W boson decays [95].

The cross-section, σ_{CC} , for an HNL particle being produced in p-p collisions via W boson production can be expressed as:

$$\sigma_{CC} = \sigma(pp \rightarrow W) \cdot B(W \rightarrow \ell N) \quad (7.3)$$

$$= \sigma(pp \rightarrow W) \cdot B(W \rightarrow \ell \nu) \sum_{\ell} |V_{\ell}|^2 \left(1 - \frac{m_N^2}{m_W^2}\right)^2 \left(1 + \frac{m_N^2}{2m_W^2}\right), \quad (7.4)$$

where $\sigma(pp \rightarrow W)$ denotes the cross-section for on-shell W boson production in p-p collisions; $B(W \rightarrow \ell N)$ is the branching fraction for a W boson decaying leptonically to an HNL particle; V_{ℓ} is the coupling strength between the HNL particle and an SM neutrino of flavour ℓ ; m_N is the mass of the HNL particle and m_W is the mass of the W boson.

7.1.1 Heavy neutral lepton signal simulation

The analysis presented in this Chapter relies heavily on HNL signal and SM background samples generated using MC techniques, as described in Sec. 5.1. These MC samples are used to determine optimal event selections and categorisations, visualise the distributions of various key observables, train and test MVA algorithms used in the search, and predict the number of expected signal and data-driven events to interpret the results. The HNL signal samples employed in this search are generated at leading-order (LO) using MADGRAPH5_aMC@NLO v2.6.5, based on the theoretical model described in Ref. [155].

A total of 76 generated signal samples were produced for both Dirac and Majorana HNL scenarios for different HNL masses in the range 1-24 GeV and different cou-

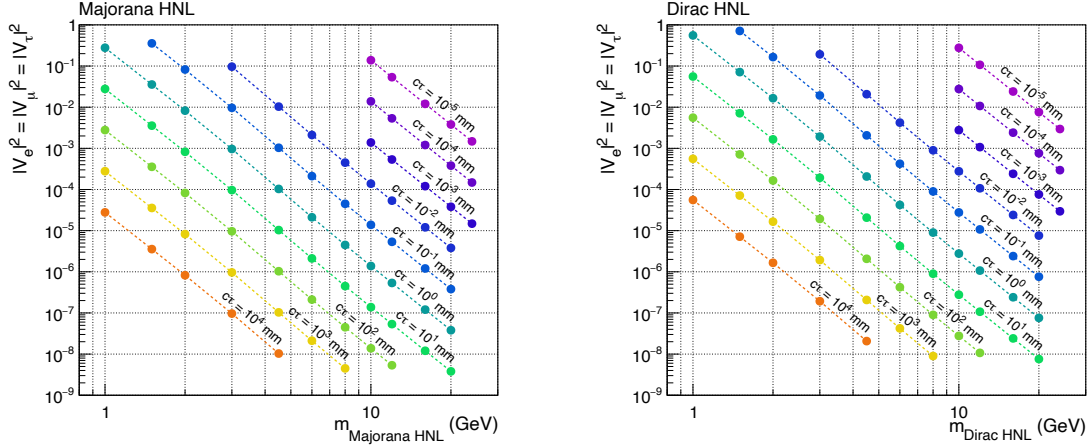


Figure 7.2: Generated HNL signal samples using MADGRAPH in the mass-coupling plane for both Majorana (left) and Dirac (right) scenarios, with equal couplings to all three lepton generations ($|V_e| = |V_\mu| = |V_\tau|$). The colour-coded, tilted lines correspond to HNLs with the same proper lifetime.

pling strength V_ℓ values, scanning HNL proper lifetimes, as presented in Fig. 7.2. For simplicity, we will refer to the mean proper decay lifetime (τ_0) or the mean proper decay length ($c\tau_0$) as just “lifetime” or “decay length”, unless stated otherwise. The samples were generated with equal coupling strengths for all three SM neutrinos, that is, $V_e = V_\mu = V_\tau$. Lepton flavour violation can occur in this HNL model if more than one coupling V_ℓ is non-zero. Simulated HNL events can be reweighted to different coupling scenarios by calculating additional matrix element (ME) weights with MADGRAPH5_aMC@NLO, thus enabling efficient scanning of the mixing space [156]. Therefore, only one signal sample needs to be generated, thus minimising computational time and memory usage.

Since the HNL mass, m_N , and proper lifetime, τ_0 , is fixed during the event generation and subsequent detector simulation, the arbitrarily reweighted coupling combinations must stay consistent with the total decay width (Γ_{tot}) of the generated HNL process:

$$\frac{1}{\tau_0} = \Gamma_{\text{tot}}(m_N, V_e, V_\mu, V_\tau) = \sum_{\ell=e,\mu,\tau} \Gamma_\ell(m_N, V_\ell) = \sum_{\ell=e,\mu,\tau} \Gamma'_\ell(m_N, V'_\ell). \quad (7.5)$$

For each generated HNL sample, 66 equidistant weights are stored for reweighting the sample to different coupling scenarios while keeping the total decay width constant. The barycentric coordinates f_e , f_μ , and f_τ are used to represent the relative ratios between the couplings $|V_e| : |V_\mu| : |V_\tau| \equiv f_e : f_\mu : f_\tau$, where $\sum_{\ell=e,\mu,\tau} f_\ell = 1$. The chosen coupling points in the f_ℓ -space are shown in Fig. 7.3.

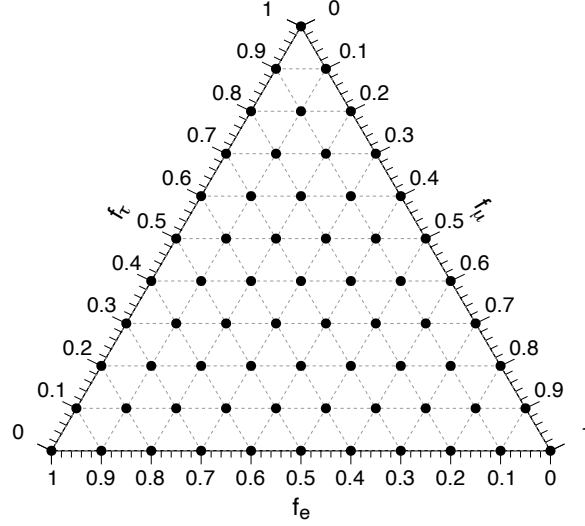


Figure 7.3: The 66 relative coupling points used for each HNL signal sample generated with equal couplings to the 3 SM neutrino generations $|V_e| = |V_\mu| = |V_\tau|$. These coupling points are expressed in barycentric coordinates ($f_e + f_\mu + f_\tau = 1$).

The HNL production cross-sections at LO were found to be compatible with the next-to-leading order (NLO) cross-sections within the estimated scale uncertainty. To normalise each LO signal cross-section, an averaged correction factor of ($k = \sigma_{\text{NLO}}/\sigma_{\text{LO}} \approx 1.10$) was applied. The LO scale uncertainty was taken as a theoretical uncertainty.

Table 7.1 presents an overview of the branching fractions for various HNL masses decaying via a range of tree-level decay modes. The branching ratios for Majorana HNLs are the same as those of the Dirac-type HNLs if the charge-conjugated decay modes are taken into account.

Table 7.1: Branching ratios calculated by MADGRAPH for the tree-level decay modes of Dirac HNLs at various masses.

Decay mode	HNL mass			
	1 GeV	5 GeV	10 GeV	20 GeV
$N \rightarrow \nu q \bar{q}$	28.5%	21.1%	18.3%	19.0%
$N \rightarrow \nu \bar{\nu} \nu$ (invisible)	12.3%	7.9%	6.1%	5.5%
$N \rightarrow \ell^- \ell^+ \nu$ (leptonic)	22.0%	22.6%	24.1%	24.2%
$N \rightarrow \ell^- q \bar{q}$ (semi-leptonic)	37.2%	48.4%	51.5%	51.4%

Fig. 7.4 presents the kinematic properties at generator-level for a variety of benchmark HNL signatures with different masses and lifetimes, that are close to the expected sensitivity limits for the search presented in this Thesis. HNLs are pro-

duced in the central pseudorapidity region, their transverse momentum distribution peaking at around 40 GeV. Generator-level kinematic properties of the leading and subleading leptons ($\ell_{1,2}$) associated with various coupling scenarios of a particular HNL benchmark model are also shown. While they are all typically produced in the central pseudorapidity region, their p_T distributions vary significantly for pure-tau coupling scenarios ($|V_e| = |V_\mu| = 0; |V_\tau| > 0$), since the majority of the energy from the decaying τ particle is carried away by undetected neutrinos. Consequently, the detectable daughter particle (either electron or muon) becomes rather soft. As a result, it is technically extremely challenging to trigger on such events, with a soft leading lepton.

The lowest possible p_T thresholds are highly desirable for ℓ_2 , regardless of its flavour. The analysis is only sensitive to leptonically decaying tau particles since trigger strategies for hadronically decaying taus with sufficiently low p_T thresholds are not available. Consequently, for pure-tau coupling scenarios, if either tau particle decays hadronically, such events would be missed as only one lepton would be produced in the event, while the analysis selection requires the presence of two.

In order to ensure a statistically significant simulated signal sample, a large number of events ($\sim 300\text{k}$) had to be generated for all chosen benchmark models and data-taking years. Consequently, a total of $300\text{k} \times 76$ (benchmark model) $\times 2$ (Dirac and Majorana) $\times 3$ (years) $\approx \mathcal{O}(100\text{M})$ MC events were generated for this analysis. Moreover, the sample size is effectively reduced by a factor of nine due to reweighting if the HNL couples only to a single lepton flavour, therefore the sample size of pure-tau events is reduced by default for reliable statistical inference. These samples were subdivided into statistically independent samples for MVA training, as well as validation studies and statistical analysis in order to avoid overtraining bias. An additional set of signal events has also been produced for validation studies of the various MVA algorithms employed in this analysis, including the BDT event classifier (see Chap. 8). These enriched signal samples contain approximately 3M individual events for eight HNL benchmark models. These events were selected using a $p_T > 20$ GeV filter at generator level.

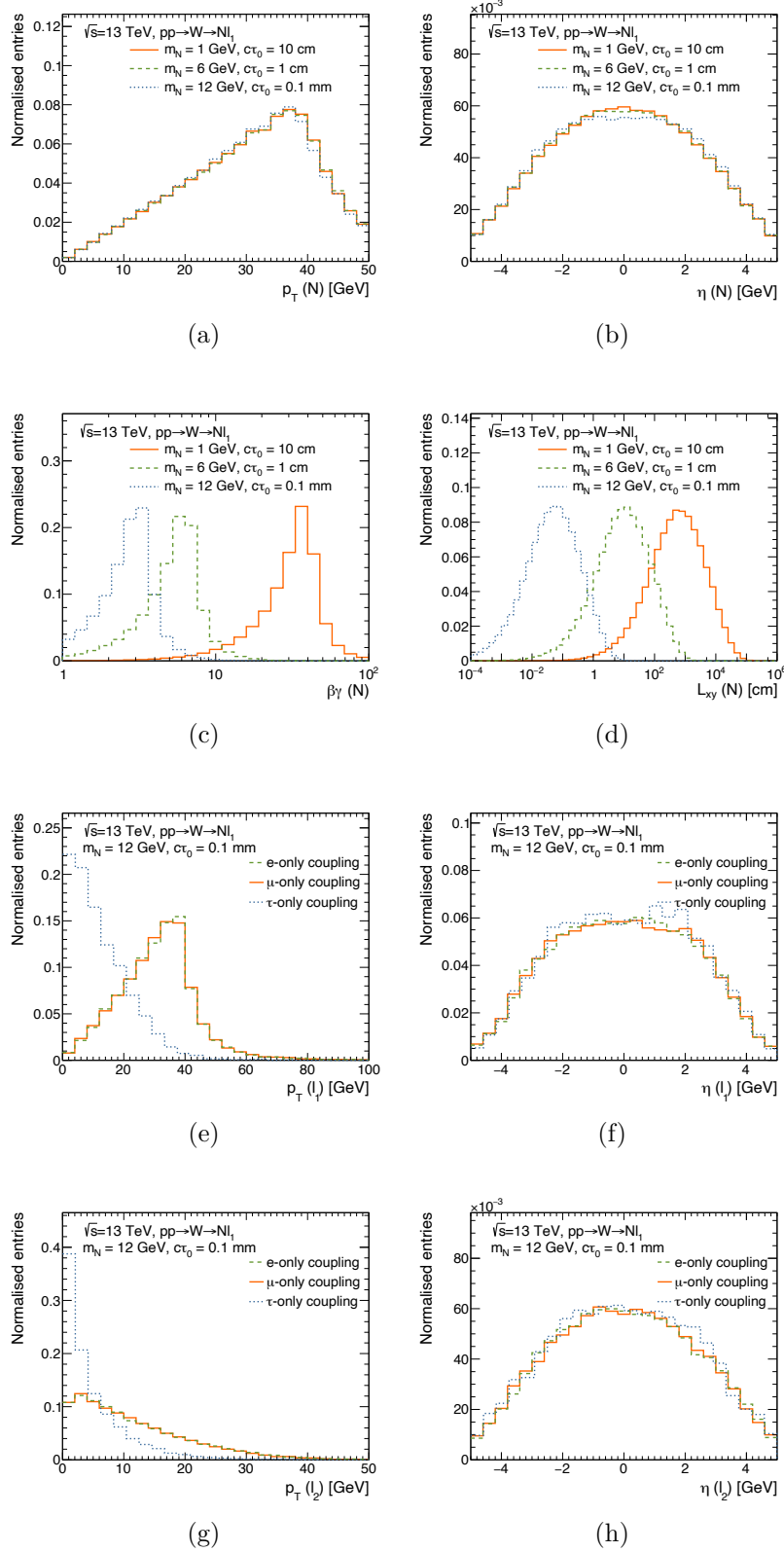


Figure 7.4: Generator-level kinematic distributions of HNL particles and associated leptons for various benchmark models and coupling scenarios: (a) HNL transverse momentum; (b) HNL pseudorapidity; (c) HNL boost; (d) HNL transverse displacement; (e) ℓ_1 transverse momentum; (f) ℓ_1 pseudorapidity; (g) ℓ_2 transverse momentum; (h) ℓ_2 pseudorapidity [109].

7.2 Background processes

Understanding the potential sources of SM background contributions and having control of their correct modelling is a crucial step in this analysis. Despite the heavy dependence on simulated SM samples, the analysis employs data-driven techniques for the estimation of the expected number of background events in the signal region (see Sec. 10.1). Consequently, the perfect modelling of all individual physics observables used in the analysis is not essential in terms of interpreting the obtained results.

As for the generated HNL signal samples, the available SM background samples were divided into two statistically independent subsets employed for different tasks within the analysis. The smaller subset was used for training the two MVA algorithms utilised for signal selection against background, whereas the larger MC background set was employed for MVA validation, data to simulation comparisons of key physics variables and statistical inference of the obtained results. In order to avoid potential overtraining bias, it is crucial to train the MVA algorithms on statistically independent data sets from those used for performance evaluation.

Preliminary selection requirements were made to maximise HNL signal yields while significantly reducing SM background contributions, resulting in the following SM processes acting as main background sources:

- **W+jets** events consist of a W boson decaying leptonically and associated ISR, FSR or pileup jets. The W decays into a prompt (leading) lepton that the event is triggered on and a neutrino, whereas an additional lepton is produced by the accompanying jets. The subleading lepton may be either a real one produced via hadronisation of heavy-flavoured jets or a "fake" lepton masked by misreconstructed or misidentified tracks. MC samples used for statistical inference were generated using MADGRAPH5_aMC@NLO at NLO [157] in the associated QCD process in the ME calculation where the parton showering (PS) of accompanying jets were simulated with PYTHIA. MC samples used for MVA training were produced with MADGRAPH at LO precision. LO samples are preferred for MVA training in order to avoid events with negative generator weights, as discussed in Sec. 5.1, whereas NLO samples provide better accuracy for the rest of the analysis.
- **Z/ γ^* +jets** events consist of prompt Drell-Yann (DY) decays to a pair of leptons ($Z/\gamma^* \rightarrow \ell\ell$) and accompanying jets. One lepton from the DY decay serves as the leading lepton, whereas the other one is a prompt subleading lepton. Such background events have significant contributions at low HNL displacement

categories, particularly the oppositely-charged same-flavour (OSSF) ones. In addition, if one of the leptons from the DY decay is lost or misidentified, lepton contributions (either real or fake) from the accompanying jets can become significant in other categories. Simulated samples used for statistical inference were produced using MADGRAPH5_aMC@NLO in the high dilepton mass region ($m_{\ell\ell} > 50$ GeV) and MADGRAPH in the low dilepton mass region ($10 < m_{\ell\ell} < 50$ GeV) at NLO and LO precision, respectively. The PS of associated jets were simulated with PYTHIA. MC samples employed for MVA training were generated at LO with MADGRAPH in the high dilepton mass region only.

- **$V\gamma^*$ +jets** events consist of a vector boson decaying leptonically in association with a prompt photon that may be misreconstructed as a jet or lepton, as well as other accompanying jets. Such backgrounds have a similar manifestation as Z/γ^* +jets events. $Z(\rightarrow \ell\ell)\gamma^*$ decays were generated with MADGRAPH at LO, whereas $W(\rightarrow \ell\nu)\gamma^*$ decays were simulated with MADGRAPH5_aMC@NLO. No $W\gamma^*$ processes are used for MVA training while a $Z\gamma^*$ sample produced via MADGRAPH5_aMC@NLO is used for training the displaced jet tagger, but not the BDT as their contribution in the signal region is relatively low and contains negative weights, which is a problem discussed in Sec. 5.1.
- **QCD multijet** events may produce one or more real leptons via decays of heavy-flavour jets, as well as contributions from fake lepton rates. Leptons rarely occur in QCD events; however, the generally high rate of QCD multijet production poses a significant background contribution in long-lived HNL searches. Nevertheless, QCD processes containing two leptons are rather challenging to simulate as they are rare compared to QCD multijet events without them, so the MC generators need to be tuned accordingly. Therefore, large generator weights can be assigned to some events, while the sample size stays statistically reduced. As a result, large fluctuations with high statistical uncertainties are observed in QCD events throughout the analysis. However, this does not have a significant effect on the final result, as the background yield is estimated using a data-driven technique. Likewise, the training of the DNN tagger does not suffer from this issue because it does not require the presence of two lepton events (see details in Sec. 7.5). On the other hand, the BDT event classifier could potentially benefit from using real data to better model QCD multijet events. QCD multijet samples were generated with PYTHIA at LO precision for both types of MC background samples employed in the search.
- **Top background** contributions from both $t\bar{t}$ and single-top events, where the top quarks decay either leptonically or semi-leptonically, resulting in a

prompt leading lepton and displaced subleading lepton(s) from b-jets. Top quark backgrounds produced using the POWHEG event generator at NLO accuracy were used for statistical inference, whereas for MVA training, $t\bar{t}$ events generated with MADGRAPH at LO were used.

The colour-coding of SM backgrounds described above corresponds to the colours used in plotting them in various control and signal region plots in the analysis.

7.3 Trigger strategy and data sets

The analysis described in this Thesis uses two sets of p-p collision data (*single muon* and *single electron/photon*) at $\sqrt{s} = 13$ TeV, collected between 2016 and 2018 by the CMS detector with single, isolated lepton triggers. The data sets correspond to integrated luminosities of 35.9 fb^{-1} , 41.5 fb^{-1} and 59.7 fb^{-1} for each data-taking year, respectively. As discussed in Sec. 7.1, the HNL signatures investigated in this analysis manifest themselves as a prompt, isolated lepton, ℓ_1 (either electron or muon) from an on-shell W boson decay, which the event is triggered on, as well as a displaced, softer secondary lepton, ℓ_2 (e or μ) and a displaced jet identified with the DNN jet tagger.

The transverse momentum thresholds applied in the single lepton trigger are shown in Table 7.2. Due to imperfect detector resolution (called *turn-on effect*) the offline leading lepton selection has a slightly increased (2 GeV) p_T threshold in order to ensure $\sim 100\%$ efficiency.

Table 7.2: p_T thresholds employed by the single lepton triggers used in the analysis for all three data-taking years. Offline p_T thresholds were increased by 2 GeV to improve the selection efficiency.

Trigger	2016	2017	2018
Single muon	$p_T > 24 \text{ GeV}$	$p_T > 27 \text{ GeV}$	$p_T > 24 \text{ GeV}$
Single electron/photon	$p_T > 27 \text{ GeV}$	$p_T > 32 \text{ GeV}$	$p_T > 32 \text{ GeV}$

Some events with multiple energetic, isolated leptons of different flavours may overlap in the two data sets used in each data-taking year. Hence, a dedicated cross-cleaning procedure is used to avoid double counting such events. Leptons are ordered by their p_T and the highest p_T lepton passing a tight isolation requirement is assigned to be the leading lepton (ℓ_1), while the second highest p_T lepton is assumed to be ℓ_2 , such that $p_T(\ell_1) > p_T(\ell_2)$. Requiring a second, loose lepton has around 38% efficiency in selecting HNL signal events, as shown in Fig. 7.8. This surprisingly low

efficiency is a result of the experimentally rather challenging low-energy signature presented by HNL decay products. In addition, any event with the presence of a third lepton is rejected, and all accepted events must have at least one reconstructed jet. The reconstructed leading leptons are geometrically matched within a $\Delta R < 0.3$ ($\Delta R < 0.1$) window to the prompt electron (muon) the event was triggered on by the HLT. In case the matching fails, the event is discarded.

Dilepton triggers could potentially further lower the p_T requirements on the leading lepton. However, such triggers have been found highly inefficient for selecting displaced leptons, requiring significantly higher p_T thresholds than the typical ℓ_2 p_T spectrum, which is rather soft (Fig. 7.4g).

7.4 Event selection and categorisation

This Section describes the main physics objects and event observables employed in this analysis, along with their associated selection criteria (known as *cuts*) based on the HNL signal event topology and trigger decision strategies used. Such event-level selection requirements restrict the phase space of possible physics processes by significantly reducing SM backgrounds while maximising HNL signal sensitivity.

Once all event-level preselection criteria have been passed, correct modelling of MC simulated events is checked against collision data in control regions (CRs) where the signal contamination is expected to be minimal. Robust control over simulation is a crucial step in not only understanding both signal and background events in the signal region (SR) but also feeding well-modelled event samples for training and validation purposes to the MVA algorithms employed in the search. Nevertheless, the final results are determined using a data-driven background estimation reducing the reliance on simulated samples.

A refined categorisation scheme ensures broad reach and interpretability of the analysis results. Event categories are chosen based on dilepton sign and flavour combinations in order to test whether the HNLs are Dirac or Majorana-type particles obeying or violating lepton flavour universality and lepton number conservation. In addition, events are categorised by the displacement of the long-lived HNLs, testing a broad range of HNL lifetimes and masses, which also impact the boost of the HNL particle. Consequently, the decay products of the HNL can be either collimated (merged) into a single displaced jet or the second lepton can be separated (resolved) from the jet associated with the HNL decay.

7.4.1 Physics objects

Primary vertex

Events with at least one well-reconstructed ("good") primary vertex (PV) are required. Good PVs are defined as those reconstructed close to the IP that have at least five high purity tracks associated with it, which also lie within $|d_z| < 24$ cm and $|d_{xy}| < 2$ cm of the beam line. The primary p-p interaction vertex is chosen to be the good PV with the highest p_T^2 obtained by summing the energy of its associated tracks. The rest of the good PVs are considered to be pileup vertices. Further explanation of vertex reconstruction in CMS is presented in Sec. 4.1.2.

It should be noted that the analysis does not explicitly require the presence of secondary vertices (SVs), in order to allow the search to include longer HNL lifetimes since highly displaced SVs are generally poorly reconstructed in CMS. In addition, this preserves sensitivity to prompt signal scenarios. Nevertheless, the set of reconstructed SVs that share at least one track with the jet is passed as inputs to the displaced jet tagger.

Muon candidates

Two types of muons are used in this analysis. The muon candidate identified as the prompt leading lepton (ℓ_1), must meet tight identification and isolation criteria [90], as discussed in Sec. 4.2.1. Consequently, the lowest p_T thresholds they pass must be consistent with the trigger selection criteria discussed in Sec. 7.3, with a slight increase in p_T in order to ensure efficient offline selection of events: $p_T > 26, 29, 26$ GeV for 2016, 2017 and 2018, respectively. The leading muon track must originate from the vicinity of the primary vertex within $|d_{xy}| < 0.01$ cm and $|d_z| < 0.05$ cm.

In contrast, muons identified as subleading leptons are only required to meet loose identification criteria. No requirement is applied to the muon isolation and its associated track is not required to be compatible with the primary vertex. Subleading muons must exceed a p_T threshold as low as 3 GeV. Both muon candidates must be within the tracker acceptance, $|\eta| < 2.4$.

Dedicated identification and reconstruction scale factors for loose muons are derived in Secs. 9.2.1 and 9.2.2. Muon object reconstruction in CMS is presented in detail in Sec. 4.2.1.

Electron candidates

As was the case with the muons, two types of electron candidates are required for this analysis. The leading electron is required to meet tight, MVA-based identification criteria [108] (see Sec. 4.2.2), as well as passing the following (offline) trigger p_T thresholds: $p_T > 29, 34, 34$ GeV for 2016, 2017 and 2018, respectively. Electron candidates found within $\Delta R < 0.05$ of a PF muon candidate with $p_T > 3$ GeV are rejected in order to avoid selecting photons radiated from muons. Prompt electrons are required to originate from close to the PV, that is, within $|d_{xy}| < 0.05$ (0.10) cm and $|d_z| < 0.10$ (0.20) cm if they pass through the barrel (endcap) system.

Subleading electron candidates must fulfil a modified version of the loose identification criteria imposed by the PF algorithm. Loose electrons are required to exceed a p_T threshold of 5 GeV. Both electron candidates must be contained within the tracker acceptance, $|\eta| < 2.4$.

Dedicated identification scale factors for loose electrons are presented in Sec. 9.3. A more detailed discussion of CMS electron object reconstruction is provided in Sec. 4.2.2.

Jets

One of the key physics objects of this analysis are displaced jets resulting from the decay of a long-lived HNL. Such jets are identified as novel deep neural network-based (DNN) displaced jet tagger, briefly introduced in Sec. 4.2.3. A general description of CMS jet reconstruction is provided in the same Section. The performance of the jet tagger and a more detailed description of its architecture are discussed in Sec. 7.5. A distinguishing feature of this analysis is that jets that contain leptons are explicitly considered, whereas, in many standard CMS analyses, such jets are always cross-cleaned with priority given to the lepton.

Jets in this analysis must satisfy the tight PF identification criteria, their calibrated transverse momentum must exceed 20 GeV and they must lie within the tracker acceptance, $|\eta| < 2.4$. The cutflow diagram shown in Fig. 7.5 presents further selection criteria for HNL candidate jets. If the jet overlaps with the isolated leading lepton ℓ_1 within $\Delta R < 0.4$, it is rejected as the lepton is the prioritised physics object. In contrast, if the subleading lepton is contained within a jet candidate, the latter may be accepted if its p_T is above 30 GeV and the lepton is not isolated ($I_{\text{rel}}^{\mu,e} > 0.15$). This ensures that the selected jets are not clustered from highly isolated leptons, as the hadronic tracks contained in the jet likely originate from pileup. Such jets are

also typically not very well-calibrated. Raising the jet p_T threshold for jets merged with leptons improves the modelling of various kinematic quantities incorporating the jet's kinematic information while maintaining sensitivity to HNL signatures. Jets separated from the subleading lepton must pass the standard 20 GeV p_T threshold.

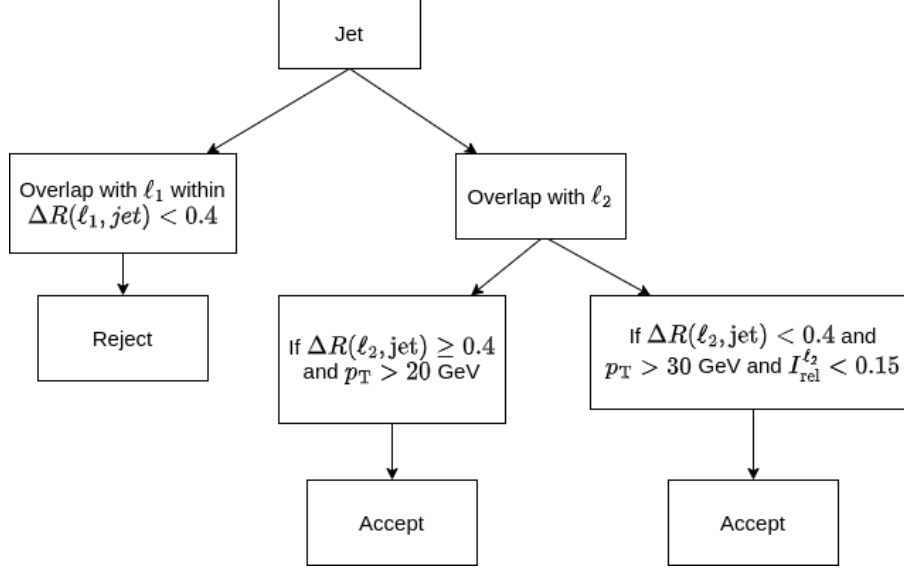


Figure 7.5: Cutflow diagram presenting the selection of jets in the analysis.

HNL jet candidates denoted j^* , are those closest in ΔR to the subleading lepton after passing the criteria listed above. HNL jets can be categorised as *merged* if they contain the ℓ_2 within $\Delta R(\ell_2, j^*) \leq 0.4$, or *resolved* if they are separated from the loose lepton, $\Delta R(\ell_2, j^*) > 0.4$, as shown in Fig. 7.6. The distribution of this variable for different benchmark models is shown in Fig. 7.7c.

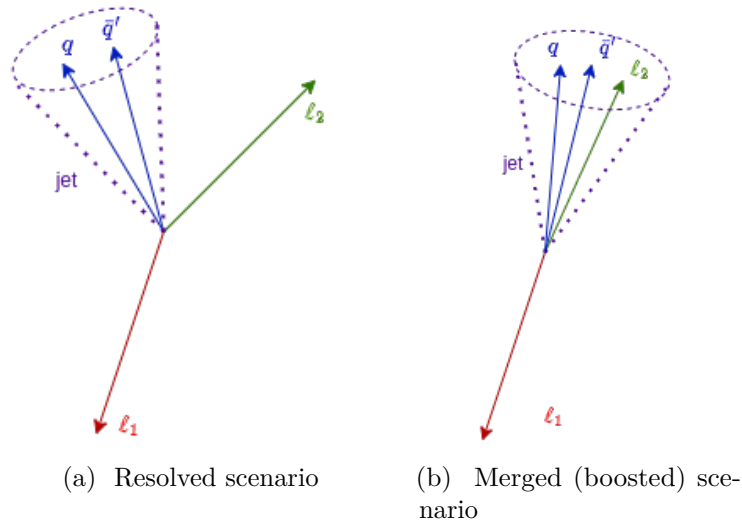


Figure 7.6: Diagrams presenting the categorisation of events by separation of the jet and the subleading lepton, ℓ_2 .

Missing transverse energy

Given the semi-leptonic topology of the final states studied in this analysis, no major missing transverse energy is expected, particularly for zero tau coupling scenarios. If $V_\tau > 0$, the missing transverse energy carried away by the neutrinos is comparable to the typical detector missing-energy resolution of about 30 GeV. A set of filters is implemented to reject events containing spurious $\vec{p}_T^{\text{miss}}$ originating from various detector effects. A description of the reconstruction of missing transverse energy in CMS is given in Sec. 4.2.4.

Other event-level observables

Further event-level observables constructed from the properties of the physics objects described above are used in defining the SR and CRs used in this analysis, as well as input variables in the training process of the event-level boosted decision tree aiming to capture prompt topologies of signal events while rejecting SM backgrounds. These observables are as follows:

- $m(\ell_1, \ell_2)$: the invariant mass of the dilepton system, used to reject events from $Z/\gamma^* + \text{jets}$ processes.
- $\Delta\phi(\ell_1, \ell_2)$: the angular separation between ℓ_1 and ℓ_2 in the transverse plane.
- $\Delta R(\ell_1, \ell_2)$: the angular separation between ℓ_1 and ℓ_2 in the $\eta - \phi$ plane.
- $\Delta\phi(\ell_1, \vec{p}_T^{\text{miss}})$: the angular separation between ℓ_1 and $\vec{p}_T^{\text{miss}}$ in the transverse plane.
- $m_T(\ell_1, \vec{p}_T^{\text{miss}})$: the transverse mass of the $\ell_1 - \vec{p}_T^{\text{miss}}$ system, used to discriminate against $W + \text{jets}$, defined as

$$m_T(\ell_1, \vec{p}_T^{\text{miss}}) = \sqrt{2 \times p_T^{\ell_1} \times \vec{p}_T^{\text{miss}} \times (1 - \cos(\Delta\phi(\vec{p}_T^{\text{miss}}, \ell_1)))} \quad (7.6)$$

- $\Delta R(\ell_2, j^*)$: the angular separation between ℓ_2 and the HNL jet candidate, denoted j^* (defined in Subsection "Jets" above), in the $\eta - \phi$ plane.
- $m(\ell_1, \ell_2, j^*)$: the 3-body invariant mass of the dilepton and j^* candidate system, defined as

$$m(\ell_1, \ell_2, j^*) = \begin{cases} \sqrt{P(\ell_1) + P(j^*)} & \text{if } \Delta R(\ell_2, j^*) \leq 0.4, \\ \sqrt{P(\ell_1) + P(\ell_2) + P(j^*)} & \text{if } \Delta R(\ell_2, j^*) > 0.4. \end{cases} \quad (7.7)$$

where $P(x)$ denotes the four-momentum of particle x . In order to avoid double counting ℓ_2 in the merged scenario, the two definitions above are used in the definition of the 3-body invariant mass variable.

Event shape angular observables

Several event shape variables are considered as inputs to the event level BDT, exploiting geometric correlations between the reconstructed physics objects in the signal event topology. These angular observables are derived from the eigenvalues of the sphericity tensor [158]:

$$S^{\alpha\beta} = \frac{\sum_i p_i^\alpha \cdot p_i^\beta}{\sum_i |\mathbf{p}_i|^2}, \quad (7.8)$$

where $\alpha, \beta = x, y, z$, and $\vec{p}_i$ is the 3-momentum of a physics object i from the collection of jets and leptons present in an event.

The eigenvalues of this tensor are equal if the event has a spherical distribution. If two or more jets are present in a collision event, at least one eigenvalue is zero.

From the eigenvalues of the sphericity tensor, $\lambda_1 > \lambda_2 > \lambda_3$, the angular observables can be calculated as follows [159]:

- Sphericity: $S = \frac{3}{2}(\lambda_2 + \lambda_3)$
- Aplanarity: $A = \frac{3}{2}(\lambda_3)$
- C-parameter: $C = 3(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)$

In addition, the event isotropy and circularity variables are used [160]:

- Isotropy: $I = \frac{\max(I_i) - \min(I_i)}{\max(I_i)}$ where $I_i = \sum_{jets} |p_x^{jet} \cos \phi_i + p_y^{jet} \sin \phi_i|$
- Circularity: $C = \min \left(\frac{\pi I_i}{2 \sum_{jets} \sqrt{p_x^{jet} + p_y^{jet}}} \right)$

where i is the i^{th} granular step in the ϕ plane. The value of the circularity and the isotropy is 1 for spherical events and 0 for linear events in the $r - \phi$ plane.

Distributions of the event shape variables listed above in the two control regions of the analysis (defined in Sec. 7.4.3) are presented in Appendix C, Figs. C.17-C.16.

7.4.2 Preselection of events

The preselection criteria listed below limit the phase space to enriched in HNL final states, with high signal acceptance and significantly reduced background from SM processes. Events with one isolated, tight lepton and one loose lepton are accepted, meeting the criteria discussed in the sections above. The leading lepton must have higher p_T than the subleading, loose lepton. Events with the presence of a third or more leptons are vetoed. At least one jet must be present in accepted events but those with more than five are rejected. Kinematic requirements on accepted jets and HNL jet candidates are presented in the Section above. Missing transverse energy filters are applied in order to avoid events with large misreconstructed missing energy.

The preselection criteria listed above are flexible enough to be employed in further selections defining the signal and control regions of the analysis.

7.4.3 Analysis selections

The signal region (SR) for this analysis lies in a phase space where HNL particles are most likely to be present. This region is kept blinded throughout the analysis, only using simulated samples to estimate the expected signal and background yields in this phase space, particularly for signal interpretation reasons once the SR gets unblinded. Therefore, it is crucial to have confidence in the quality of the simulated samples used throughout the analysis. As a result, two control regions (CRs) have been defined by inverting some of the selections applied for the SR, hence minimising signal contamination while maintaining the freedom to study the SM backgrounds in data and simulation. This subsection presents a brief description of the selection criteria used to define the SR and the two CRs, which are also summarised in Table 7.3. Figure 7.7 shows the distributions of kinematic variables for the different signal hypotheses used in the definition of the preselection criteria and analysis selection requirements.

Table 7.3: Overview of the preselection requirements, as well as definitions of the signal and control regions for the analysis. The kinematic selection criteria for tight leptons are shown per data-taking year (2016-2018).

	Signal region	Drell-Yann (DY) CR	Resolved CR
Lepton selection			
Tight, isolated lepton (ℓ_1)			
Muon: $p_T > 26, 29, 26$ GeV, $ \eta < 2.4$	✓	✓	✓
Electron: $p_T > 29, 34, 34$ GeV, $ \eta < 2.4$			
Loose lepton (ℓ_2)			
Muon: $p_T > 3$ GeV, $ \eta < 2.4$	✓	✓	✓
Electron: $p_T > 5$ GeV, $ \eta < 2.4$			
$p_T(\ell_1) > p_T(\ell_2)$	✓	✓	✓
Third lepton veto	✓	✓	✓
Jet selection			
Merged jets ($p_T > 30$ GeV, $I_{\text{rel}}^{\ell_2} > 0.15$, $ \eta < 2.4$)	✓	✓	✓
Resolved jets ($p_T > 20$ GeV, $ \eta < 2.4$)	✓	✓	✓
$1 \leq N_{\text{jet}} < 5$	✓	✓	✓
Event observables			
$p_T^{\text{miss}} < 100$ GeV	✓	✓	✓
p_T^{miss} filter	✓	✓	✓
$\min \Delta R(\ell_2, \text{jets}) < 1.3$	✓	✓	inverted
$m(\ell\ell) > 20$ GeV	✓	✓	✓
$m(\ell\ell) < 80$ GeV	✓	inverted	✓
$70 \text{ GeV} < m(\ell_1, \ell_2, j^*) < 90 \text{ GeV}$	✓		

Signal region

The key variable defining the SR is the minimal opening angle between the subleading lepton and the HNL jet candidate, $\min \Delta R(\ell_2, \text{jets})$. Since the HNL products are expected to be collimated (as shown in Fig. 7.7c), $\min \Delta R(\ell_2, \text{jets}) < 1.3$ is required. An upper limit of 100 GeV is applied on the p_T^{miss} to reduce the contribution from W+jets events with significant missing energy. In addition, further selection requirements are applied to the dilepton mass of the event, based on Fig. 7.7d, such that signal events are expected to lie between $20 < m(\ell_1, \ell_2) < 80$ GeV. The lower threshold is expected to suppress QCD multijet events, while the upper bound removes $Z/\gamma^* + \text{jets}$ background, which is significant in opposite-sign, same-flavour (OSSF) event categories. Fig. 7.8 shows a cutflow graph depicting the efficiency of signal event selection after each cut is applied for a benchmark HNL process and two main SM background processes, as well as data. The series of selection requirements significantly reduce rates in the background processes while maintaining high efficiency for signal selection.

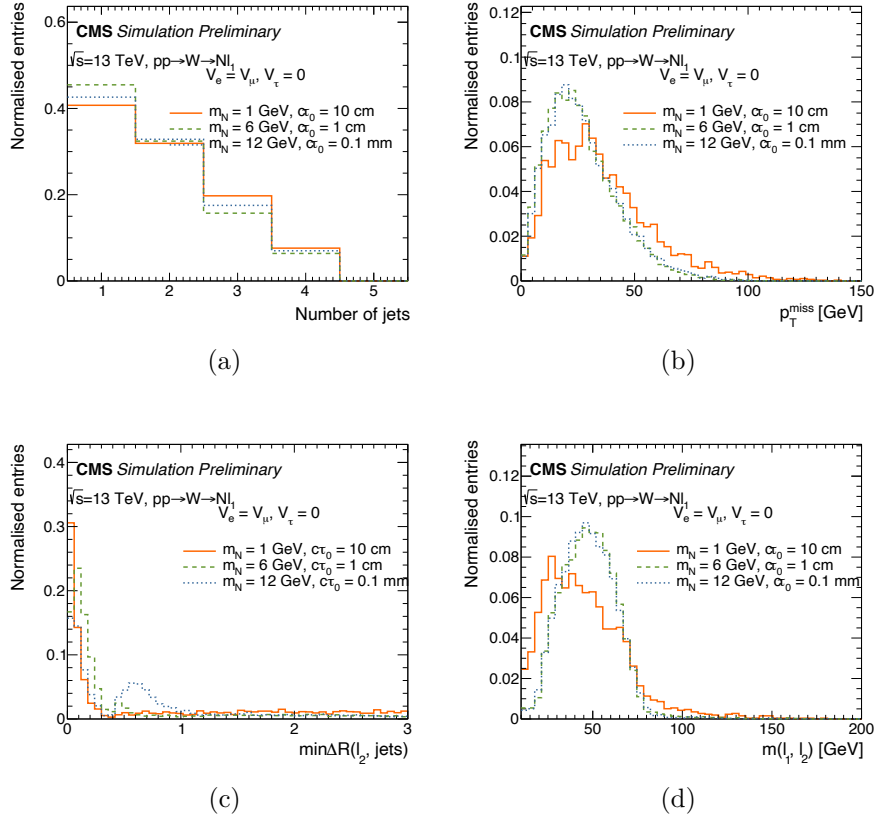


Figure 7.7: Reconstructed event observables for various benchmark HNL scenarios assuming equal coupling to the first two generations, motivating the selection criteria applied in the preselection of the analysis, as well as SR and CRs [109].

Control regions

The two CRs used in this analysis are defined by inverting one specific variable defining the SR while keeping the rest the same. This ensures minimal signal contamination in the phase space of the CRs.

The *Drell-Yann* (DY) CR is defined by inverting the upper limit on the dilepton mass ($m(\ell_1, \ell_2) > 80$ GeV), creating a CR dominated by $Z/\gamma^* + \text{jets}$, particularly in the OSSF event category.

The other CR used in the analysis inverts the ΔR opening angle selection to accept events where the subleading lepton is highly separated spatially from the HNL jet candidate: $\min \Delta R(\ell_2, \text{jets}) > 1.3$. The CR is denoted *resolved CR* and has a rather similar background composition to the SR.

The modelling of key physics object properties was inspected per year and category in both the DY and resolved CRs via data to MC comparison plots (see Figs. 7.9-7.12 for the DY CR and further validation plots, for both DY and resolved control

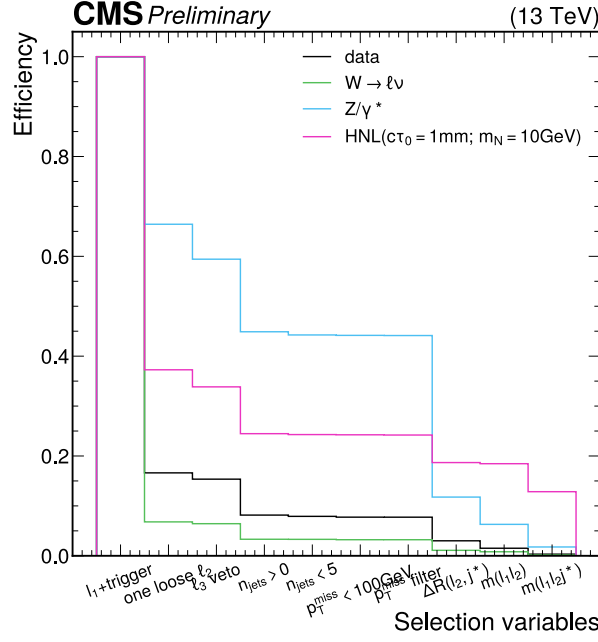


Figure 7.8: Cumulative cutflow diagram presenting event-level preselection and SR cuts after the trigger selection was applied for a representative HNL signal model (Majorana-type of $m_N = 10\text{GeV}$, $c\tau_0 = 1\text{mm}$), the two main SM background processes and data (2016).

regions with additional categories in Appendix C). Both statistical and systematic uncertainties resulting from the displaced lepton scale factors described in Chapter 9 are shown in the comparison plots. OSSF categories have $Z/\gamma^* + \text{jets}$ as the dominant background, whereas it is negligible in same-sign mixed flavour (SSMF) categories. Overall, good agreement is observed for the main analysis physics objects – jets, leptons and missing energy. A slight excess in data is nonetheless observed for low- p_T jets. Due to the limited number of events available in the QCD samples, large fluctuations are observed for this process in most of the CR distributions. Some p_T^{miss} mismodelling can be seen in 2018 data; however, this is expected when taking into account the jet energy correction uncertainties studied in Ref. [161].

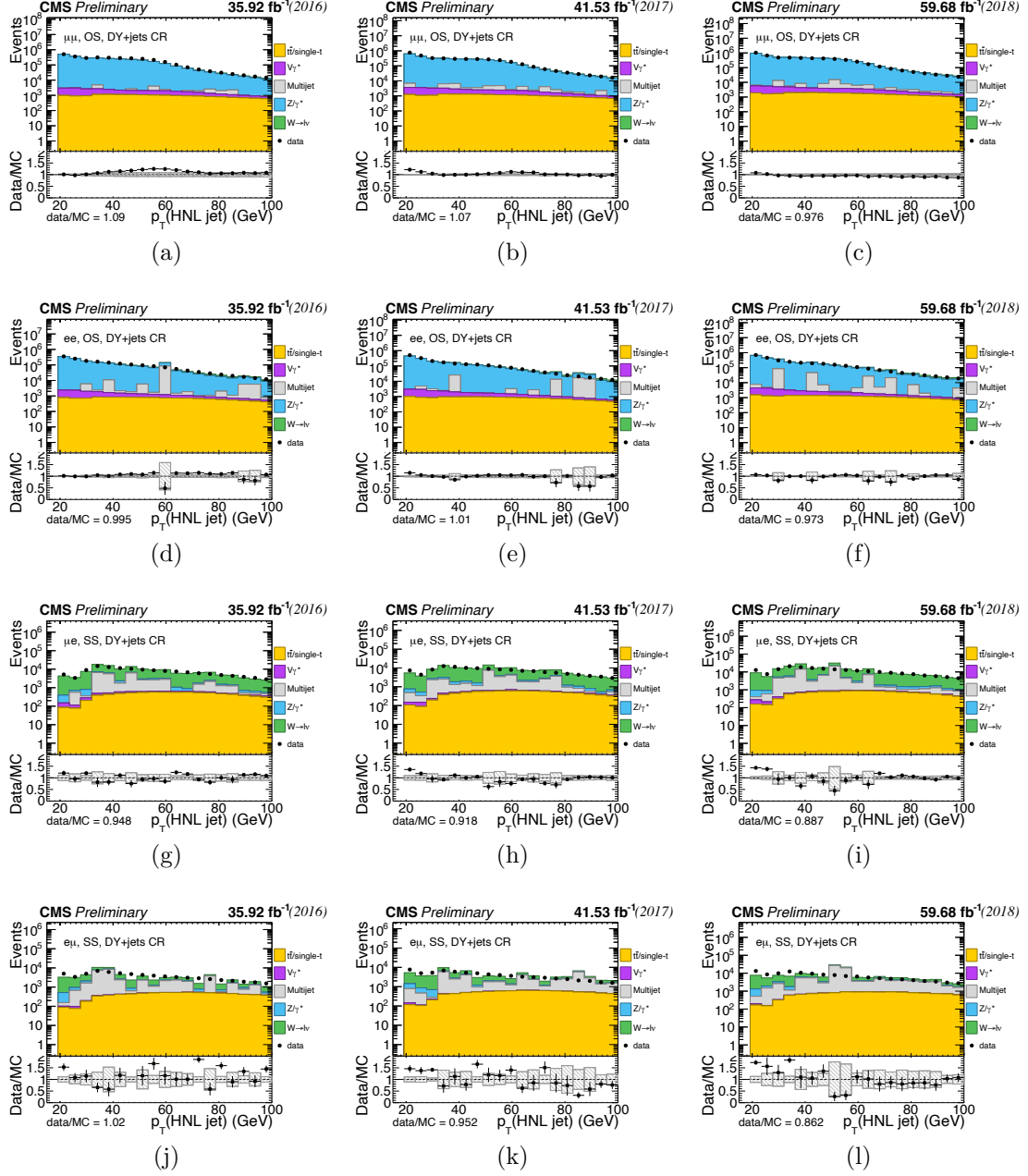


Figure 7.9: Distribution of the HNL candidate jet p_T in the DY control region. The distributions are shown for the 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

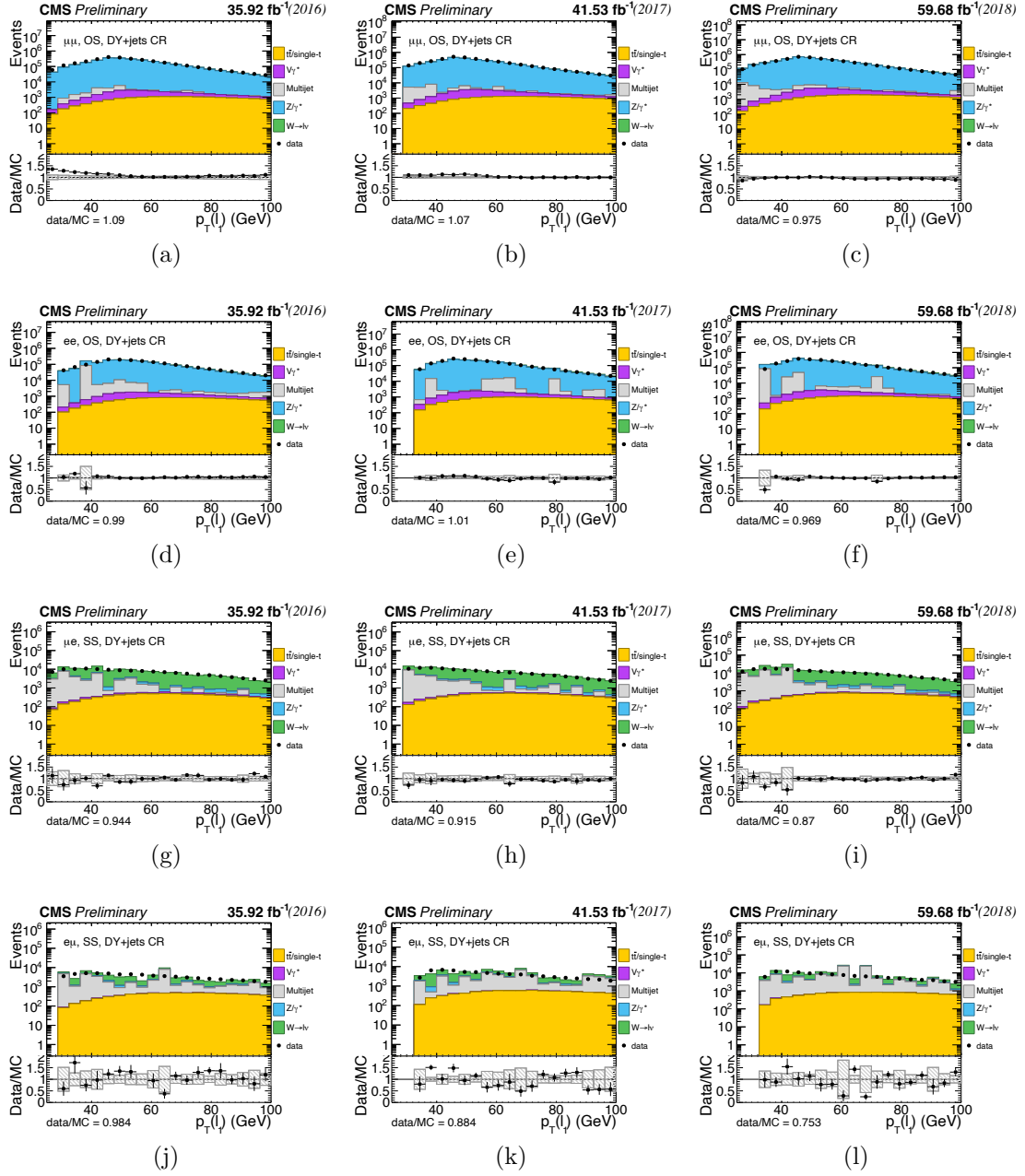


Figure 7.10: Distribution of the ℓ_1 transverse momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

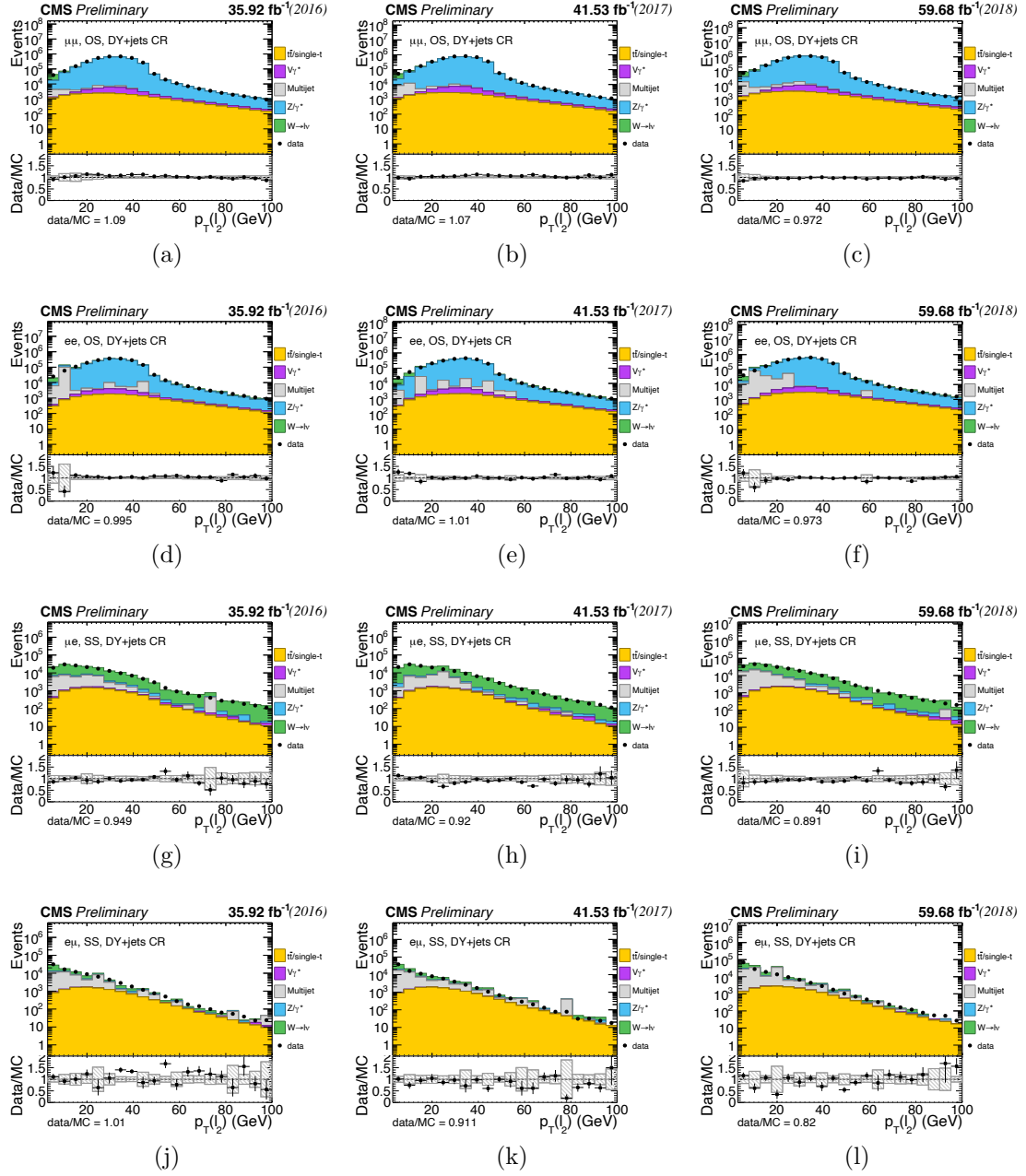


Figure 7.11: Distribution of the ℓ_2 transverse momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

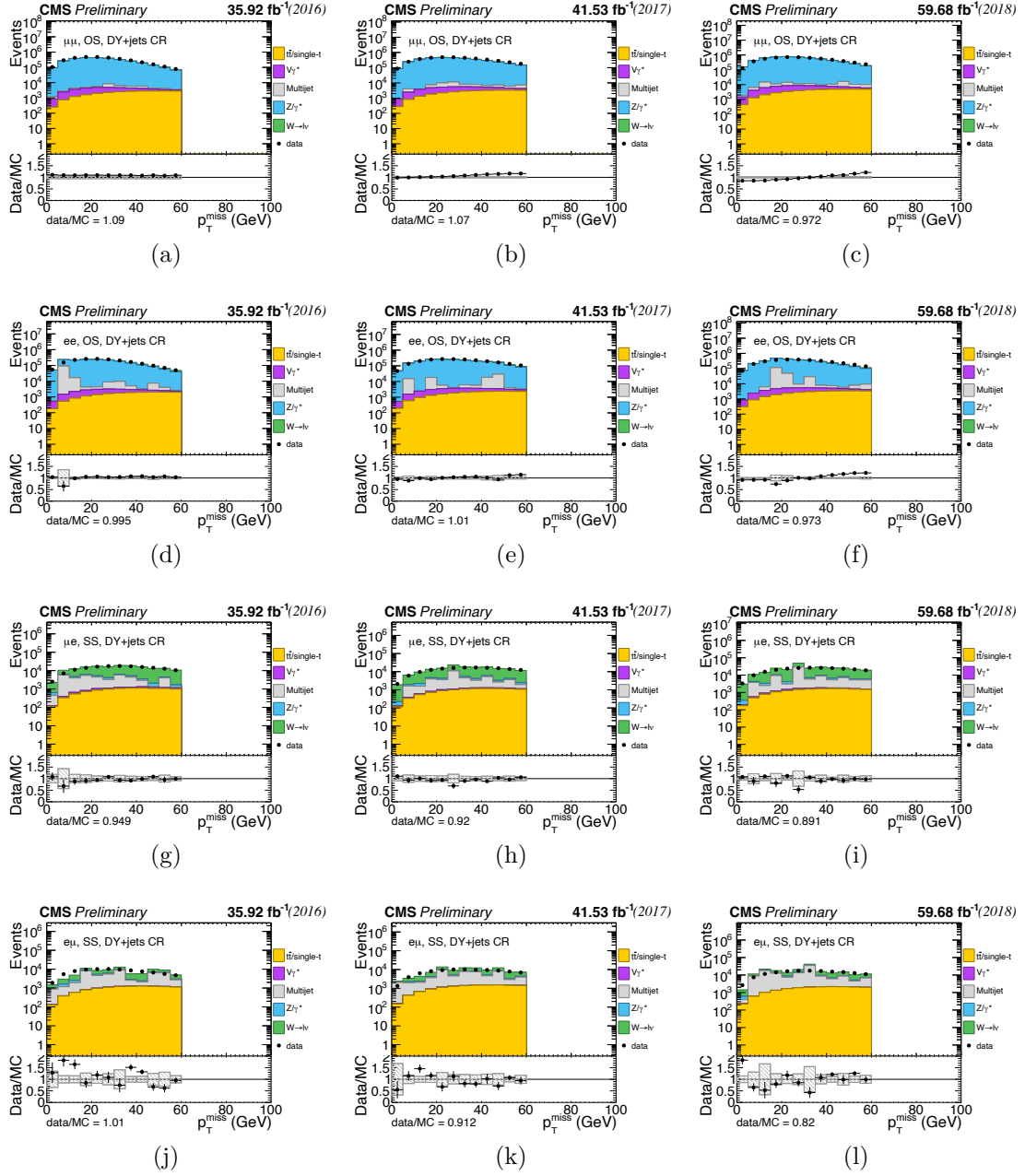


Figure 7.12: Distribution of p_T^{miss} in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

7.4.4 Event categorisation

Events in this analysis are categorised by the dilepton flavour and sign, the displacement of the subleading lepton and the separation between the HNL candidate jet and ℓ_2 . This categorisation scheme for SR events ensures high sensitivity for the different HNL benchmark processes, spanning a range of HNL mass and lifetime scenarios, as

well as different HNL mixings. A total of $4(\text{dilepton flavour}) \times 2(\text{dilepton charge}) \times 3(\ell_2 \text{ displacement}) \times 2(\text{separation}) = 48$ event categories are defined in the analysis:

- **dilepton flavour:** $\mu\mu$, μe , ee , $e\mu$, probing different HNL-SM neutrino coupling scenarios;
- **dilepton charge:** *opposite sign (OS)* and *same sign (SS)*, testing whether the HNLs are Dirac (OS only) or Majorana-type (both OS and SS) particles;
- **ℓ_2 displacement** defined by the transverse impact parameter significance variable, d_{xy}^{sig} , in order to separate prompt and displaced HNL decays. The $d_{xy}^{\text{sig}}(\ell_2)$ displacement is divided into three categories:
 - **prompt:** $d_{xy}^{\text{sig}}(\ell_2) < 1$;
 - **intermediate:** $1 < d_{xy}^{\text{sig}}(\ell_2) < 10$;
 - **displaced:** $d_{xy}^{\text{sig}}(\ell_2) > 10$;
- **ΔR separation between ℓ_2 and j^* :** SR events are separated according to whether the subleading lepton is *merged* inside the HNL candidate jet or *resolved* from it:
 - **merged:** $\min\Delta R(\ell_2, j^*) < 0.4$. Dominated by low-mass HNL processes typically have such topologies;
 - **resolved:** $0.4 < \min\Delta R(\ell_2, j^*) < 1.3$, being populated primarily by higher-mass HNL scenarios.

7.4.5 Corrections to data and simulation

Several corrections are applied to physics objects in simulation in order to enhance the agreement between data and simulation, ensuring robust control over the use of MC-simulated events. These corrections are also considered sources of systematic uncertainty, further detailed in Chapter 9. Corrections are applied to the following physics objects:

- **Jet energy resolution (JER):** The energy of the jets in the simulation is smeared such that the resolution matches the one found in data. For jets that can be linked to particle-level jets, the difference in energy between the two jets is either reduced or increased using scaling factors based on data-to-simulation studies. If a reconstructed jet cannot be matched to a particle-level jet (e.g. it originates from pileup), its jet energy is randomly increased using measured resolutions provided by the CMS collaboration [112, 114, 115].

- **Jet energy scale (JES):** The momentum of the reconstructed jets is calibrated in both data and simulation through offset and response corrections [113]. Furthermore, data-only residual corrections are applied. These corrections are also applied to p_T^{miss} .
- **Pileup:** Simulated events are reweighted in order to match the topology of pileup distributions in data, separately for each data-taking year. Corrections from data are based on the instantaneous luminosity per data-taking year and a minimum bias cross-section of 69 mb [111].
- **Muons:** Prompt muon candidate trigger, isolation and identification scale factors are applied based on measurements [90]. In the case of displaced muon candidates, there are reconstruction and identification scale factors to be considered. First, a set of muon reconstruction scale factors are applied ensuring that the selected tracks have matching segments in the muon stations, derived in Sec. 9.2.2. Then, loose identification scale factors are applied to the displaced muon candidates, derived in Sec. 9.2.1. Consequently, the overall efficiency for displaced muons is evaluated as: $\varepsilon_{\text{total}} = \varepsilon_{\text{reco}} \times \varepsilon_{\text{ID}}$.
- **Electrons:** Prompt electron identification and reconstruction scale factors are applied based on measurements, as well as displaced electron reconstruction efficiency SFs [108]. Displaced electron identification scale factors are derived in Sec. 9.3 and consequently applied in the analysis.

7.5 Deep neural network-based displaced jet tagger for heavy neutral lepton searches

A key feature of the HNL analysis is the implementation of a sophisticated deep neural network-based (DNN) displaced jet tagger designed to identify the displaced signatures of long-lived HNL decay products. The displaced jet tagger is a modification of the previously published DNN multiclassifier described in Ref. [9], including various extensions to address the experimental signature of displaced jets originating from semi-leptonically decaying HNLs. This Section provides a brief introduction to the signal and SM background jet truth classification schemes used through the training of the displaced jet tagger, as well as the DNN architecture and its validation in the analysis CRs. The output score of the displaced jet tagger is used as a discriminating variable in the data-driven background estimation (see Sec. 10.1), alongside the event-level BDT classifier presented in Chapter 8.

7.5.1 Displaced signal jet subclasses

The categorisation of the signal jets the DNN is trained on is based on the semi-leptonic HNL decay topologies. Typically, low-mass HNL decay products manifest boosted kinematics, therefore, they are often collimated (*merged*) into a single jet. If the displaced jet contains the subleading lepton originating from the long-lived HNL decay, the jet is labelled by the flavour of the ℓ_2 : **QMU** and **QE**. If the subleading lepton is separate (*resolved*) from the displaced jet, the jets are labelled as **Q**. All hadronic activity incorporated into a **Q** label jet consists of light-flavour jet constituents. No displaced heavy-flavour jet constituents are considered for signal jet labelling during training, since they are kinematically suppressed, and the coupling of HNLs to such quarks is extremely small.

Other displaced signal jet categories exist in the default DNN jet tagger, such as single lepton jets (**MU**, **E**, **TAU**), where only a signal lepton is clustered into a jet by the anti- k_T algorithm without any signal quarks. Their frequency is relatively low compared to the categories introduced above. In addition, hadronic tau decay signatures are merged into the displaced quark jet, **Q** category, as the jet tagger was found to struggle to differentiate between the two subcategories.

7.5.2 Standard model background jet classes

A number of SM background jets are considered in the training and evaluation of the DNN multiclassifier, including light-flavour (uds), gluon (g) and heavy flavour jets (c, b). In addition, pileup jets are a significant background source due to their kinematically soft signatures ($p_T < 50$ GeV), similar to those of HNL products. Prompt lepton (e, μ, τ) and photon (γ) jets are typically prompt leptons and photons clustered with spurious hadronic activity into a jet by the anti- k_T algorithm, similarly to the single lepton jets signal classes described above (**MU**, **E**, **TAU**). These SM background jets constitute a significant background for merged signal jet categories (**QMU**, **QE**). The main characteristic of prompt lepton/photon jets is the fact that an unusually large fraction ($> 50\%$) of the jet energy comes from the corresponding lepton/photon, as they do not originate from the hadronisation of the quark or gluon initiating the jet. Typically, analyses remove such jets via *overlap removal*; however, this procedure is not possible for the signal jets considered in this analysis, as jets merged with the displaced lepton are used.

7.5.3 Deep neural network architecture

The architecture of the DNN tagger was modified to accommodate displaced leptons inside the jets. Consequently, a new set of input features was added, that relate to the kinematics of the subleading lepton. Fig. 7.13 shows a schematic diagram of the DNN architecture implemented in the displaced jet tagger algorithm, including charged and neutral PF constituents, global jet features, PF muon and electron information, as well as the reconstructed SVs. However, not all events contain reconstructed SVs, therefore their existence is not an explicit requirement in the event selection of the analysis, extending the phase-space which the search covers. If no SVs are reconstructed in the jet, the impact parameter of the charged PF candidates associated with the HNL jet candidate can still be used as a discriminating feature.

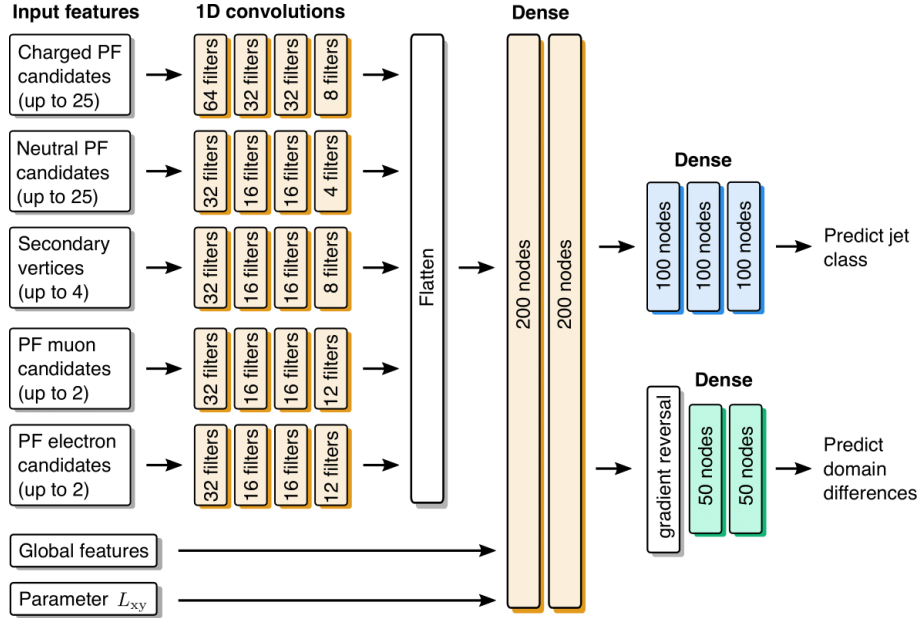


Figure 7.13: Schematic diagram of the DNN-based displaced jet tagger architecture with the goal to capture the decay products of long-lived HNLs searched in this analysis.

Besides training the DNN tagger on signal jets from the SR of the analysis, domain adaptation (DA) is used to train the DNN on jets obtained from data in the CRs in order to avoid the tagger learning differences between simulation and real data. The advantage of DA is the lack of need to fine tune the extensive set (~ 700) of low-level tagger inputs needed to ensure good data-MC agreement.

The DNN jet tagger is parametrised by the transverse displacement, L_{xy} , of the HNL particle obtained from generator-level information, in order to have a single network for arbitray displacements. For the case of background jets, where such displacement

information is not available, the L_{xy} is sampled from the signal distribution [162].

A profile likelihood ratio is used to evaluate the tagger output for each signal class over all possible SM background classes, as the true L_{xy} is unknown. It is defined as follows:

$$P(X/\text{SM} | L_{xy}) = \frac{P(X | L_{xy})}{\sum_c^{\text{SM}} P(c | L_{xy})} \quad (7.9)$$

$$\Rightarrow \hat{P}(X/\text{SM}) = \max_{L_{xy}} [P(X/\text{SM} | L_{xy})] \text{ for } X \in \{\text{LLP_Q} + \text{LLP_QTAU}, \text{LLP_QMU}, \text{LLP_QE}\} \quad (7.10)$$

where $\hat{P}(X/\text{SM})$ is the maximum likelihood ratio of the displaced jet class over the transverse displacement. The maximum likelihood ratio discriminant is normalised from a $[0, \infty)$ range to a $[0, 1]$ range to simplify the interpretation of the tagger output. The `tagger_score` variable is the combination of the three HNL signal jet classes given the topology of an event and the flavour of the subleading lepton ℓ_2 , used in the statistical evaluation of the results:

$$\text{tagger_score} = \begin{cases} \text{Q} & \text{if } \Delta R(\ell_2, j^*) > 0.4 \\ \text{QE} & \text{if } \Delta R(\ell_2, j^*) < 0.4 \text{ \& } l_2 = e \\ \text{QMU} & \text{if } \Delta R(\ell_2, j^*) < 0.4 \text{ \& } l_2 = \mu \end{cases} \quad (7.11)$$

7.5.4 Tagger performance and validation

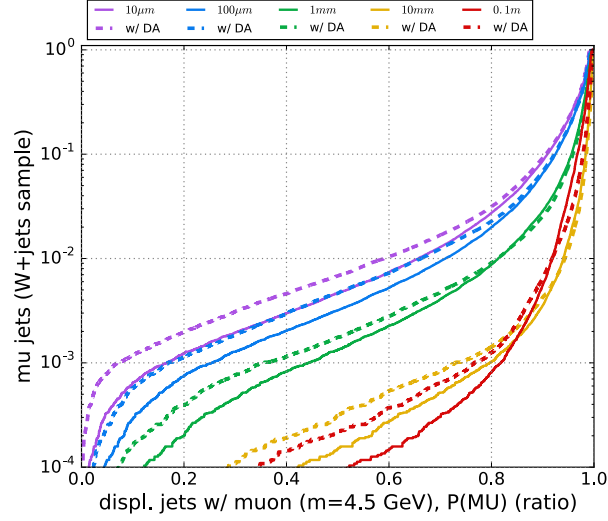
An extensive set of validation studies was performed to gain confidence in the robustness of the DNN displaced jet tagger. A number of ROC curves were constructed (see Fig. 7.14) showing that the discriminating power between the HNL decay products and SM backgrounds increases with displacement up to about 10 mm for nearly all displaced jet classes. Beyond that displacement, the discrimination power plateaus or decreases as no further tracking information is available. Only LLP_QMU jets exhibit a further increase in discrimination power up to 100 mm due to the inclusion of information from the muon system that helps identifying displaced tracks. The tagger has slightly better performance for merged HNL scenarios with respect to resolved jets. The implementation of DA did not dramatically reduce the performance of the tagger.

The tagger output was also validated through a large number of data-MC comparisons in the DY CR for both resolved (Q, Fig. 7.15) and merged (QMU and QE) scenarios (Figs. 7.16 and 7.17) in two representative dilepton categories, providing confidence in the performance of the DNN tagger. No data-MC validation distributions showing the tagger output were made in the resolved CR as merged jets cannot be selected by definition in this CR.

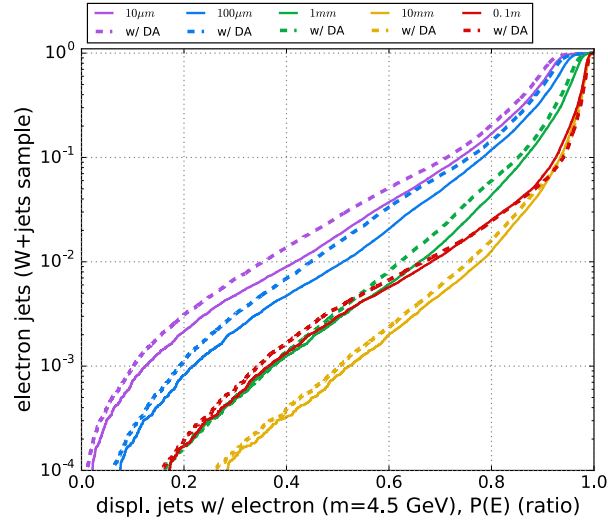
7.6 Summary

This Chapter presented an introduction to the analysis scope and strategies employed in an inclusive search for long-lived HNLs coupling to multiple generations of SM neutrinos, testing lepton flavour universality and lepton number conservation rules in the SM. The long-lived HNLs investigated are produced via CC weak decays to a prompt lepton, which the events are triggered on, and an HNL that eventually decays to a secondary lepton and a jet, marked by displaced topologies, captured via a DNN-based displaced jet tagger.

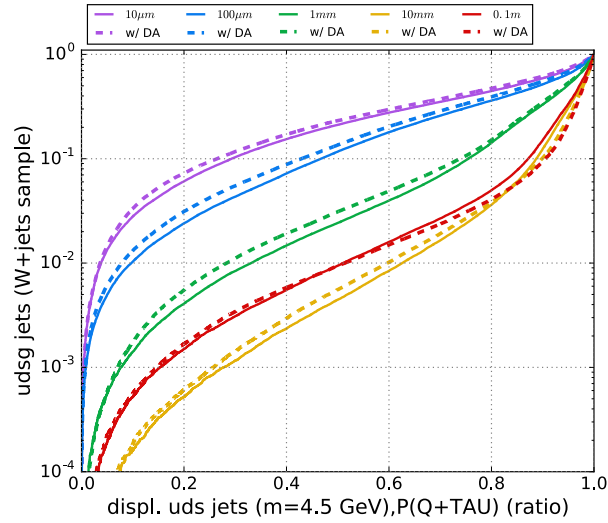
An event-level binary classifier developed to capture the prompt kinematics of signal events is the subject of the next Chapter 8, followed by a study of the sources of systematic uncertainty arising from displaced leptons in Chapter 9. The results of this search are presented in Chapter 10, based on data-driven background estimation techniques, extracting background and signal yields in the blinded SR.



(a) muon-jets



(b) electron-jets



(c) quark-jets

Figure 7.14: Tagger ROC curves for signal (a) muon-jets; (b) electron-jets and (c) quark-jets versus the equivalent SM background jet class taken from an exclusive W+jets production sample, measured for an HNL with $m_N = 4.5$ GeV with various proper lifetimes.

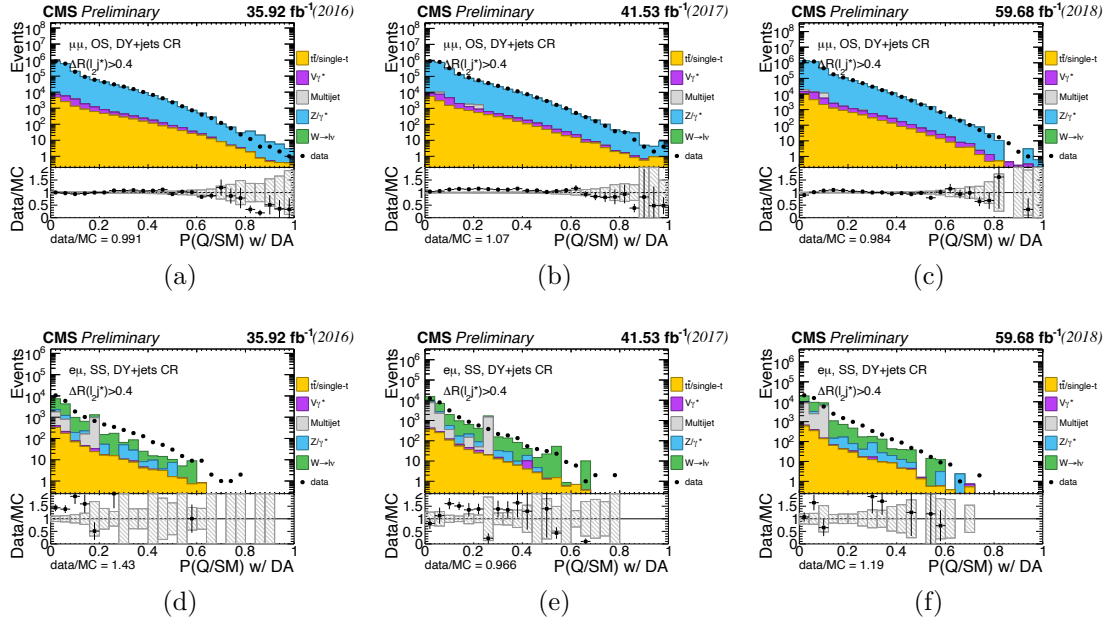


Figure 7.15: Modelling of the tagger output in the DY control region, resolved scenario. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OS dimuon category in the first row, and the SS electron-muon mixed flavour category in the bottom row. The poor agreement in the SS category results from the lack of QCD multijet events, which are one of the main background sources for this category.

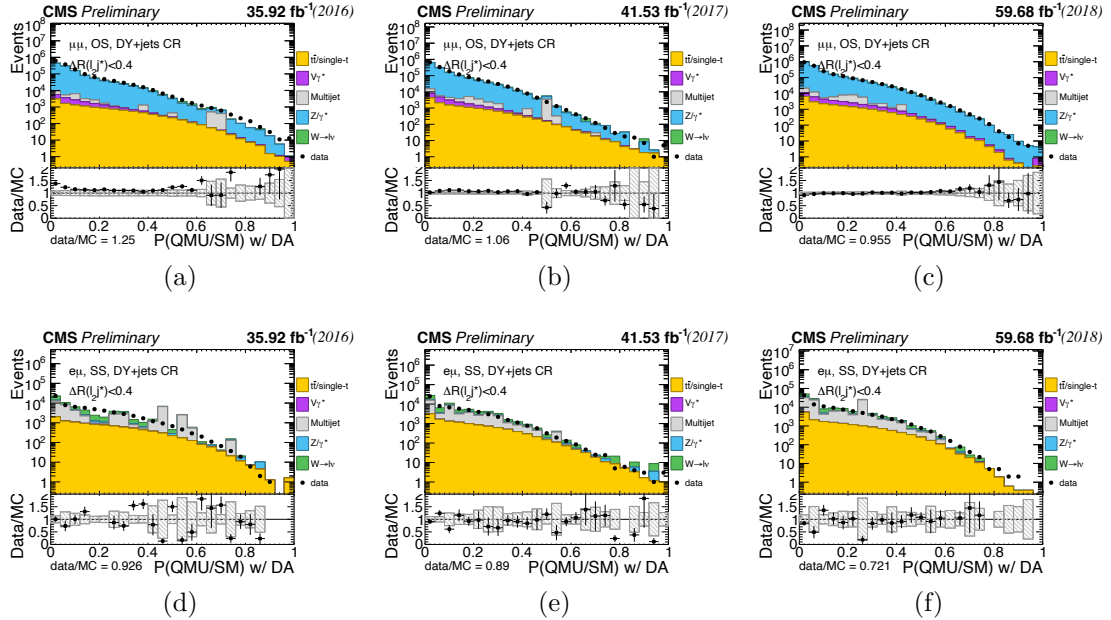


Figure 7.16: Modelling of the tagger output in the DY control region, merged muon scenario. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OS dimuon category in the first row, and the SS electron-muon mixed flavour category in the bottom row.

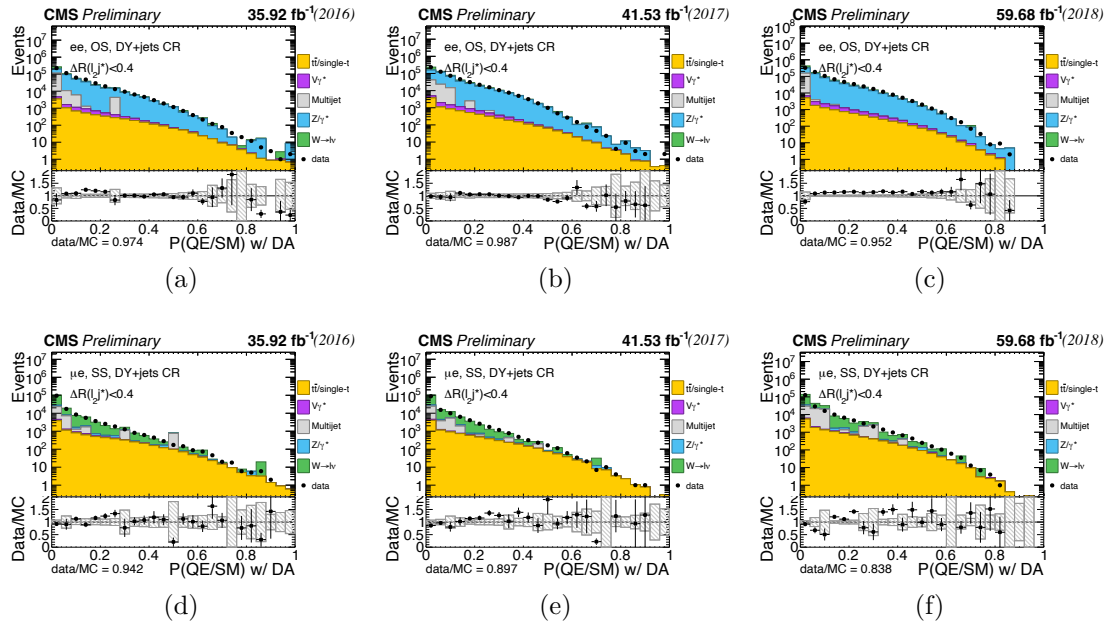


Figure 7.17: Modelling of the tagger output in the DY control region, merged electron scenario. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OS dielectron category in the first row, and the SS muon-electron mixed flavour category in the bottom row.

Chapter 8

Event-level classifier for the long-lived heavy neutral lepton search

In this Chapter, a binary event classifier employing a boosted decision tree (BDT) algorithm to separate potential HNL signal events from SM background is described. The event classifier is designed to complement the displaced jet tagger, discussed in Sec. 7.5. The jet tagger captures the "displaced features" of the HNL jet, while the BDT focuses on high-level event kinematic variables, which are expected to be independent of the displaced jet features. The event classifier output is used as a discriminating variable in the data-driven background estimation discussed in Chapter 10.

8.1 Introduction

This Section introduces the training setup, MC samples and the various input features used in the design of an optimal BDT model as an event-level discriminant.

The BDT is trained in the SR of the analysis defined in Sec. 7.4. The BDT modelling is validated in both the DY and resolved CRs introduced in Sec. 7.4. The BDT is trained on an independent set of Majorana HNL events, providing a balanced set of opposite-sign (OS) and same-sign (SS) signal dilepton events.

The first step was an exploratory study to select the relevant BDT input features. Eight dilepton event categories were employed in these studies, depending on the flavour of the dilepton system ($\ell_{1,2} = \mu, e$), as well as the dilepton system charge (i.e.,

OS & SS final states), consistent with the categorisation scheme used for the analysis (Sec. 7.4). The HNL search is sensitive to scenarios where a leptonically decaying τ particle produces a subleading lepton, ℓ_2 . In this case, the lepton was expected to have a softer p_T spectrum and the presence of missing energy from its daughter neutrinos. In contrast, other kinematic quantities in the event were expected to be unchanged.

The BDT event classifier was trained on an independent set of SM MC background samples generated at LO, presented in Table 8.1. The SM background processes used for the BDT training are consistent with those introduced in Sec. 7.2: leptonic decays of the W boson; $t\bar{t}$ /single-top production and dilepton production via the Drell-Yan process. Other background types with relevant yield of events are omitted from the training process due to low event yields in the available MC samples (e.g. multi-jet events produced via QCD processes) or lack of availability of an independent LO MC sample as in the case of vector boson fusion with γ . The classifier performance is tested on the set of NLO MC samples introduced in Sec. 7.2, employed in the event selection definition studies and background estimation of the analysis. The BDT event classifier was trained inclusively for both boosted and resolved event categories, as well as all displacement scenarios, however, a separate BDT classifier is trained for each year of data-taking.

Before training the BDT, the correlations between the potential input features were investigated. The following kinematic input features were considered as potential BDT input variables: p_T^{miss} , H_T , $|\Delta\phi(\ell_1, \vec{p}_T^{\text{miss}})|$, $m_T(\ell_1, \vec{p}_T^{\text{miss}})$, $m(\ell_1, \ell_2)$, $|\Delta\phi(\ell_1, \ell_2)|$, $\Delta R(\ell_1, \ell_2)$, $m(\ell_1, \ell_2, j^*)$, as defined in Sec. 7.4, as well as those listed below:

- dilepton charge: -1 for OS and +1 for SS;
- $p_T(\ell_1), p_T(\ell_2)$ & $p_T(j^*)$: the transverse momenta of the leading and sub-leading leptons, as well as the jet tagged as the HNL jet candidate, j^* ;
- $|\eta(\ell_1)|$, $|\eta(\ell_2)|$ & $|\eta(j^*)|$: the absolute value of the pseudorapidity of the leading and sub-leading leptons, as well as the jet tagged as HNL jet candidate, j^* ;
- the flavour of the leading and subleading lepton;
- angular variables describing the event shape: isotropy (I), circularity (C), sphericity (S), aplanarity (A) and C -parameter (defined in Sec. 7.4).

The absolute value of the opening angle and of the pseudorapidity are used as they are symmetric about the detector's central axis. Hence, no additional information

Table 8.1: SM MC background samples used for training the BDT event classifier, in each data-taking year, respectively. The table presents the number of events ($\#$ evts), the fraction of positive-weighted events (f_{pos}) and the corresponding cross-sections used for normalisation (σ).

Sample	$\#$ evts (M)	f_{pos}	σ (pb)
2016			
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8-ext1	6.9	100%	5941
TTJets_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9.6	100%	832
WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	57	100%	61526
2017			
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8-ext1	8.4	100%	5941
TTJets_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	8.0	100%	832
WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	45	100%	61526
2018			
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8-ext1	8.1	100%	5941
TTJets_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	10	100%	832
WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	70	100%	61526

can be gained from by the sign of those variables. However, taking the absolute value statistically doubles the available distributions for the positive values, and the number of events present at BDT nodes, splitting over those specific variables within the decision tree.

8.2 Correlation studies

Choosing statistically independent BDT input features minimizes the usage of redundant input variables. One quantitative method used to determine the extent of the linear correlation between any given pair of variables is the Pearson correlation coefficient, ρ , introduced in Sec. 5.3.1.

Figs. 8.1 and 8.2 present the correlation matrices of signal and background for the input features listed above. Potential linear (anti)correlation is observed between the following feature pairs in the case of signal processes, ordered by (anti-)correlation strength, $\rho > 0.4$:

- event isotropy & event circularity;
- event sphericity & event shape C;
- $|\Delta\phi(\ell_1, \ell_2)|$ & $\Delta R(\ell_1, \ell_2)$;
- $|\eta(\ell_2)|$ & $|\eta(j^*)|$;
- $p_T(\ell_2)$ & $m(\ell_1, \ell_2)$;
- event isotropy & event aplanarity;
- event circularity & event aplanarity;
- H_T & $|\Delta\phi(\ell_1, \ell_2)|$;
- H_T & $\Delta R(\ell_1, \ell_2)$;
- H_T & $p_T(j^*)$;
- H_T & event isotropy;
- H_T & event circularity;
- $|\Delta\phi(\ell_1, \ell_2)|$ & event isotropy;
- $|\Delta\phi(\ell_1, \ell_2)|$ & event circularity;

- $\Delta R(\ell_1, \ell_2)$ & event isotropy;
- $\Delta R(\ell_1, \ell_2)$ & event circularity;
- $p_T(j^*)$ & $m(\ell_1, \ell_2, j^*)$;
- p_T^{miss} & $m_T(\ell_1, \vec{p}_T^{\text{miss}})$;
- event sphericity & event aplanarity;
- event aplanarity & event shape C.

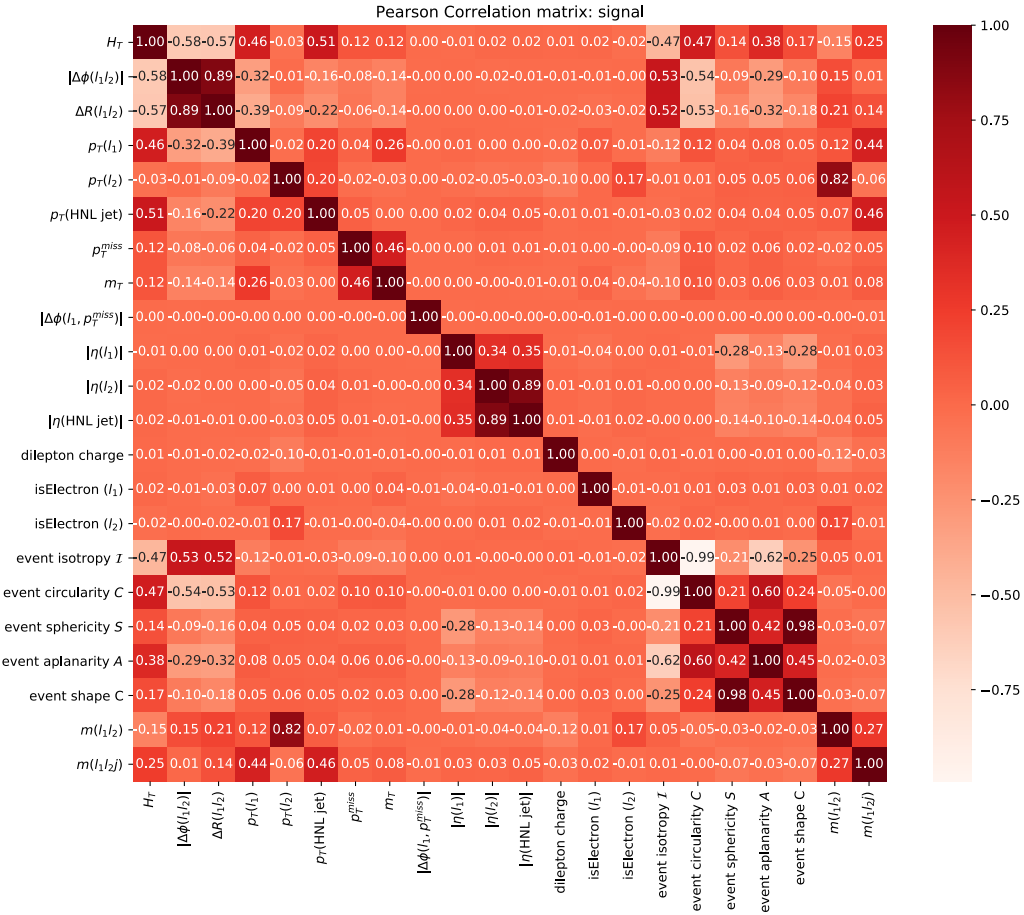


Figure 8.1: Correlation matrix presenting the linear (anti)correlation strength between all possible BDT input feature pairs. The matrix is based on the signal simulation in the SR for the 2016 data-taking period.

Potential linear correlation is observed between the following feature pairs in the case of background simulation, ordered by (anti-)correlation strength, $\rho > 0.35$:

- event isotropy & event circularity;

- event sphericity & event shape C;
- $|\eta(\ell_2)|$ & $|\eta(j^*)|$;
- $|\Delta\phi(\ell_1, \ell_2)|$ & $\Delta R(\ell_1, \ell_2)$;
- $p_T(\ell_2)$ & $m(\ell_1, \ell_2)$;
- H_T & $p_T(j^*)$;
- $p_T(j^*)$ & $m(\ell_1, \ell_2, j^*)$;
- p_T^{miss} & $m_T(\ell_1, \vec{p}_T^{\text{miss}})$;
- $p_T(\ell_1)$ & $m(\ell_1, \ell_2, j^*)$;
- event aplanarity & event shape C;
- H_T & $m(\ell_1, \ell_2, j^*)$;
- $|\Delta\phi(\ell_1, \ell_2)|$ & event isotropy/circularity;
- $\Delta R(\ell_1, \ell_2)$ & event isotropy/circularity;
- $\Delta R(\ell_1, \ell_2)$ & $m(\ell_1, \ell_2)$;
- $\Delta R(\ell_1, \ell_2)$ & $m(\ell_1, \ell_2, j^*)$;
- $|\eta(\ell_1)|$ & event sphericity;
- $|\eta(\ell_1)|$ & event shape C;
- event isotropy & event aplanarity;
- event circularity & event aplanarity;
- event sphericity & event aplanarity;
- $m(\ell_1, \ell_2)$ & $m(\ell_1, \ell_2, j^*)$.

The first two correlations are due to H_T being calculated from the $|p_T|$ of all jets in the event. There is a correlation in η between the HNL jet candidate, j^* and the subleading lepton, arising from the definition of the former. The dilepton mass is a partial input to the $m(\ell_1, \ell_2, j^*)$ variable. p_T^{miss} is a component of the $m_T(\ell_1, \vec{p}_T^{\text{miss}})$ and $|\Delta\phi(\ell_1, \vec{p}_T^{\text{miss}})|$ variables, hence the potential correlation. Correlations observed between $|\Delta\phi(\ell_1, \ell_2)|$ and $\Delta R(\ell_1, \ell_2)$, as well as, various other angular variables occur by definition.

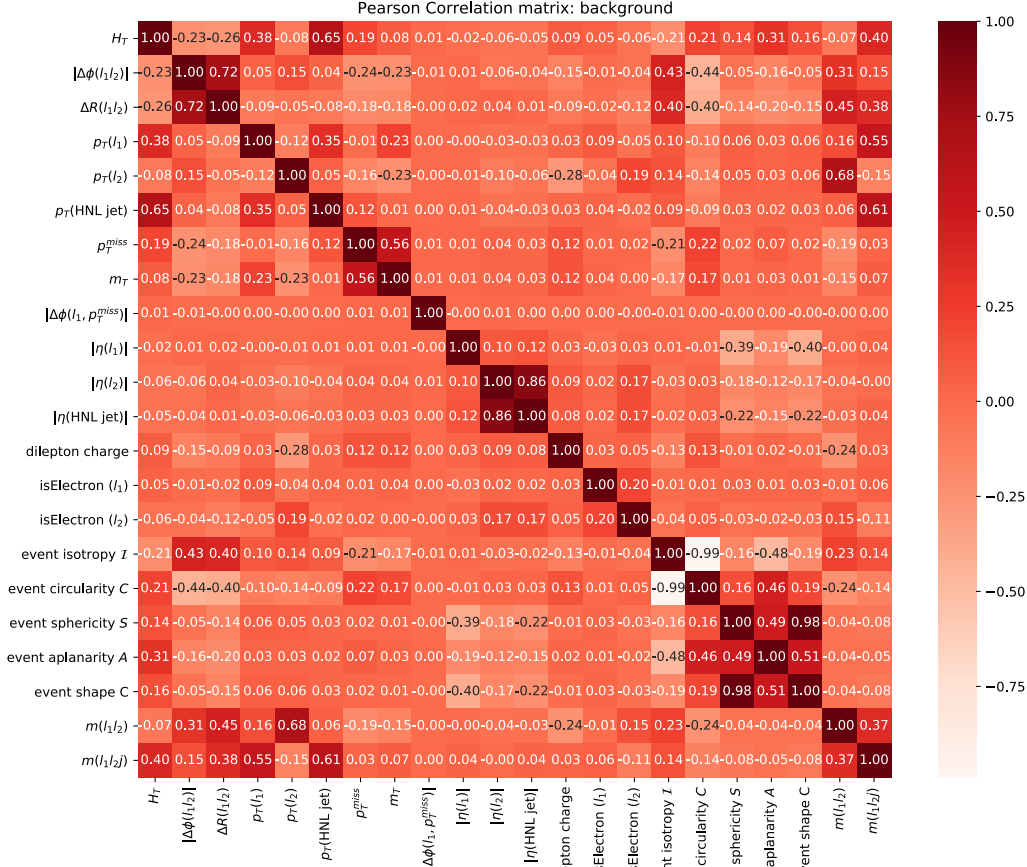


Figure 8.2: Correlation matrix presenting the linear (anti)correlation strength between all possible BDT input feature pairs. The matrix is based on the background simulation in the SR for the 2016 data-taking period.

The mutual information (MI) is a measure employed to quantify the higher-order dependence of the two variables, as introduced in Sec. 5.3.2. In this study, the normalised version of the MI is used to understand the extent of the correlation between the potential BDT input variables (Eq. 5.13).

Figs. 8.3-8.4 show a pair of correlation matrices between potential BDT inputs employing the MI metric, one for signal and one for background. The higher-order correlations observed between variable pairs are consistent with those observed with the Pearson linear correlation coefficient.

As mentioned previously, the BDT event classifier is used in the data-driven background estimation method described in Sec. 10.1. The $m(\ell_1, \ell_2, j^*)$ variable is used for event preselection before conducting the ABCD background estimation in the blinded signal region employing the displaced jet tagger information and the BDT score. The event classifier output must be fully decorrelated with the DNN jet tagger

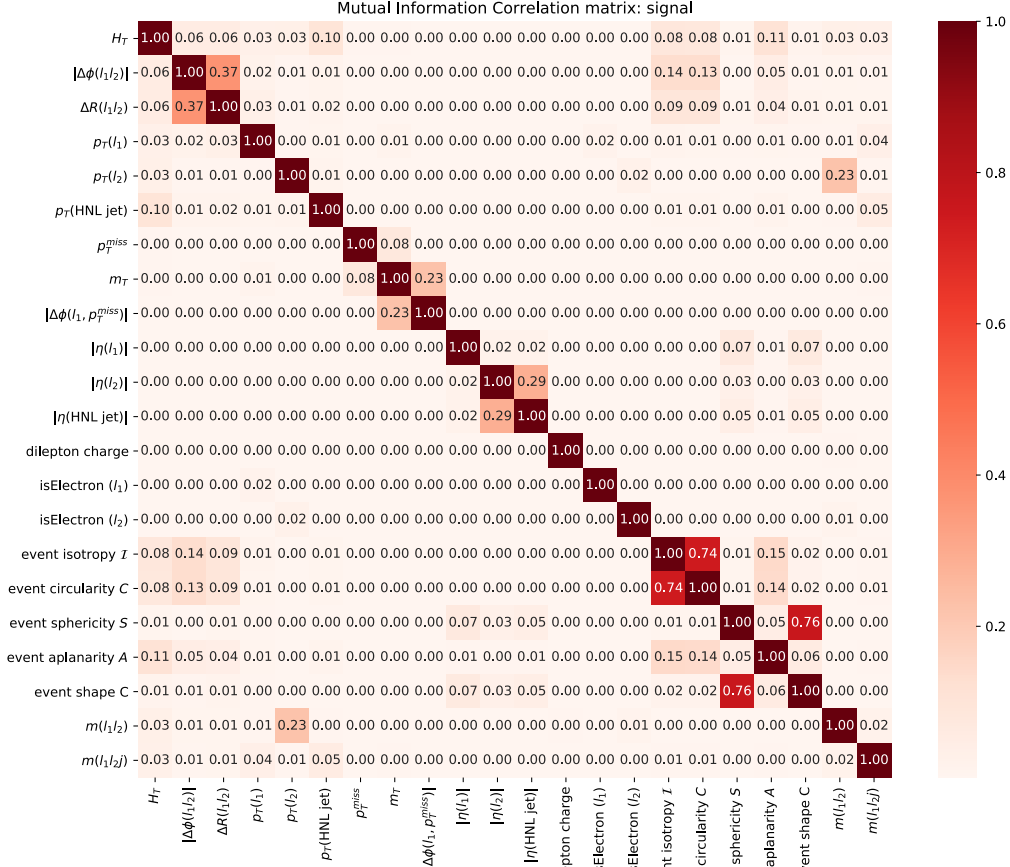


Figure 8.3: Correlation matrix, presenting the higher-order correlation strength between all possible BDT input feature pairs. The matrix is based on the signal simulation in the SR for the 2016 data-taking period.

output since the variables employed in the ABCD method must be independent. Consequently, the input variables chosen to train the event-level BDT classifier cannot be used as input variables to train the tagger. Furthermore, the BDT output is constructed to be uncorrelated with the $m(\ell_1, \ell_2, j^*)$ variable in order to minimize bias in the methods. Nevertheless, two BDT models were explored, and eventually, the *uncorrelated BDT* was implemented in the data-driven background estimation, where the BDT output is uncorrelated with $m(\ell_1, \ell_2, j^*)$ and the DNN jet tagger information variable. For reference, plots from the *correlated BDT* are presented in Appendix F.4 to prove that better BDT performance could be achieved if more input variables are used, but with the BDT score becoming increasingly correlated with the DNN tagger output.

The tagger score was evaluated depending on the event topology:

- In the merged dilepton scenario, the profiled QL tagger score was used, where L

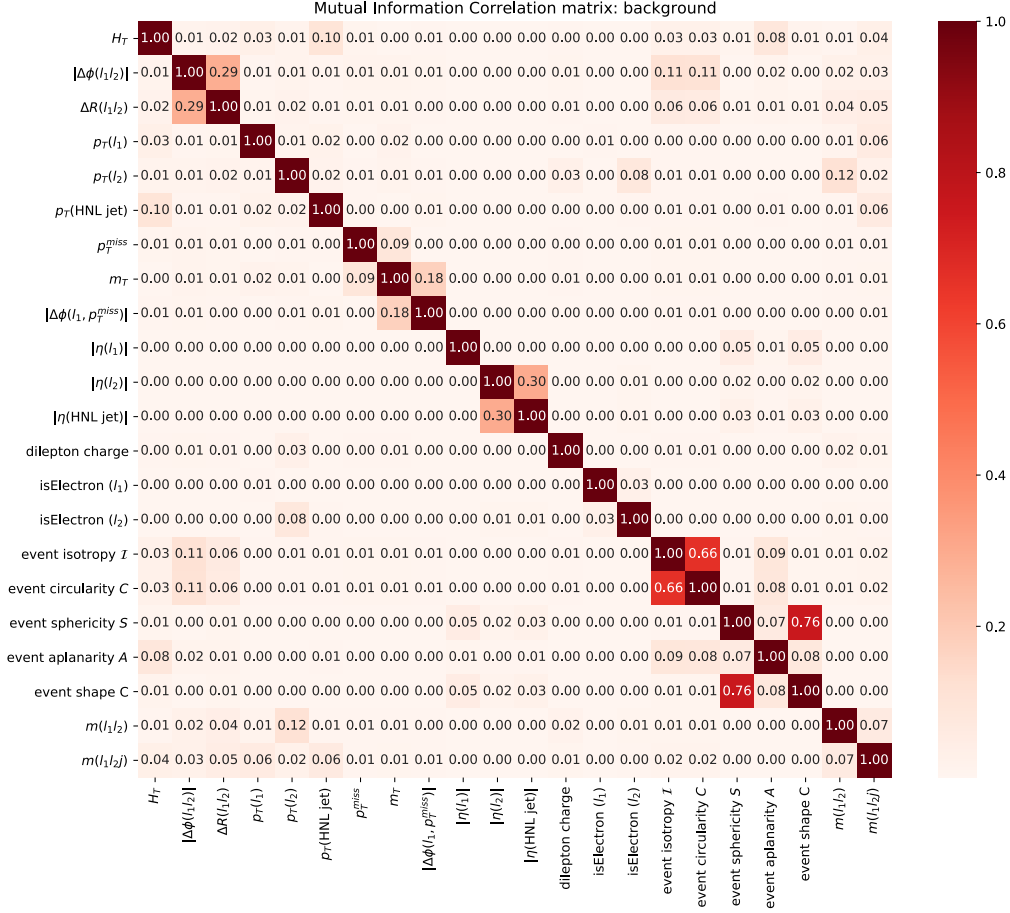


Figure 8.4: Correlation matrix, presenting the higher-order correlation strength between all BDT input feature pairs. The matrix is based on the background simulation in the SR for the 2016 data-taking period.

is the flavour of ℓ_2 .

- In the resolved dilepton scenario, the profiled Q tagger score was used.

A detailed description of the profiled tagger score evaluation can be found in Sec. 7.5.

Using an extensive set of heatmap distributions in Appendix F.1 between the tagger score and various BDT input features, the extent of any correlations can be determined. Stronger correlation is observed in the general shape and mutual information value between the tagger score and $m(\ell_1, \ell_2, j^*)$, $m(\ell_1, \ell_2)$, H_T , $p_T(\ell_2)$, $p_T(j^*)$ and $|\eta(j^*)|$. The displaced jet tagger is designed employing information about the tagged HNL jet and subleading lepton (in the merged scenario), hence the correlation observed with the above variables.

In Figs. 8.5 & 8.6, rankings of potential BDT input features by the extent of their correlation with the tagger score in the four dilepton flavour categories (dilepton

charge combined, 2016 simulation) are constructed using the Pearson correlation coefficient and the mutual information as quantitative metrics. Both metrics prove that the $p_T(\ell_2)$, $p_T(j^*)$, $|\eta(j^*)|$, $m(\ell_1, \ell_2, j^*)$, $m(\ell_1, \ell_2)$, H_T and $\Delta\phi(\ell_1, \ell_2)$ are the most correlated variables with the tagger score, providing information on building BDT models that are uncorrelated from the tagger, as well as avoiding the usage of redundant input variables (further information in Sec. 8.4). A similar set of studies was performed to investigate the correlation between $m(\ell_1, \ell_2, j^*)$ used as a preselection variable and the possible BDT input features, see Appendix F.3.

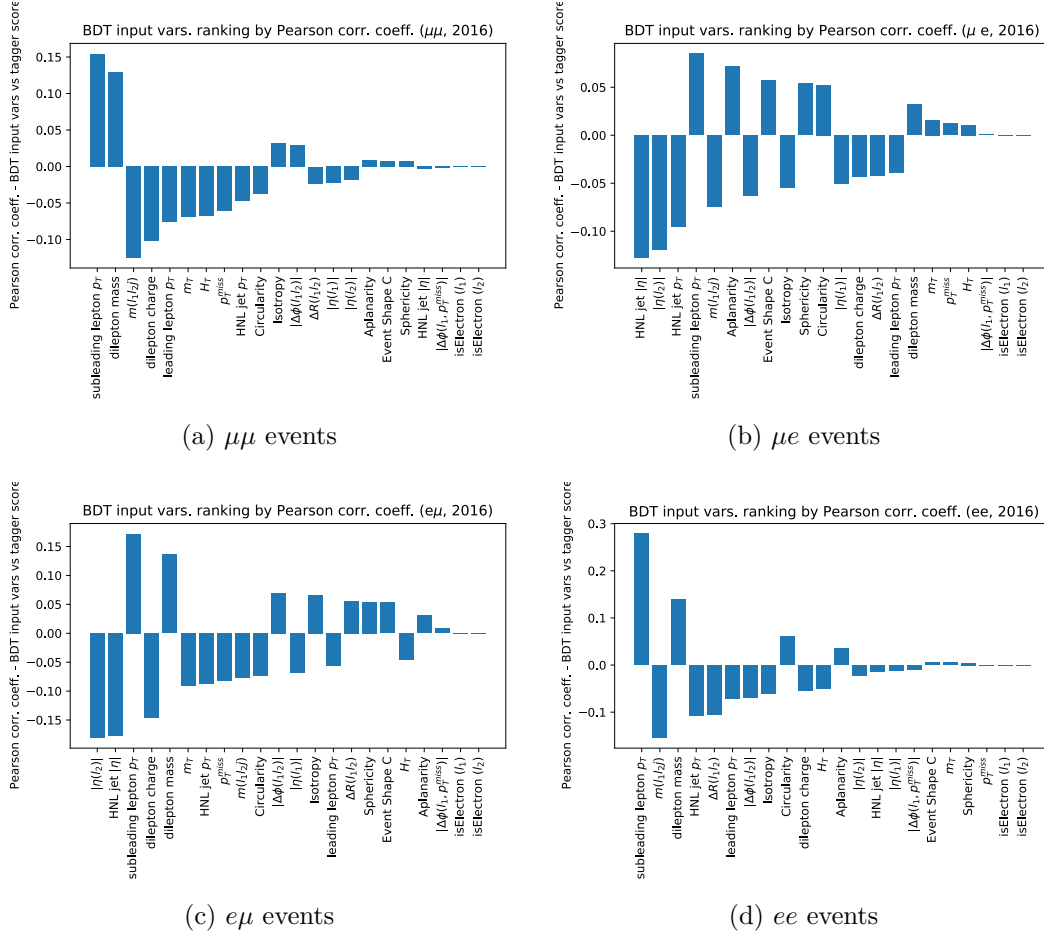


Figure 8.5: BDT input feature ranking by the extent of correlation with the tagger score, using the Pearson correlation coefficient (2016 simulation).

8.3 Data resampling

A separate set of training samples was isolated from the rest of the analysis to train the BDT event classifier and the displaced jet tagger in order to keep them unbiased during their validation studies. The training simulation set contains $\sim 20\%$ of each generated Majorana HNL sample ($\sim 85,000$ events) described in Sec. 7.1. A subset

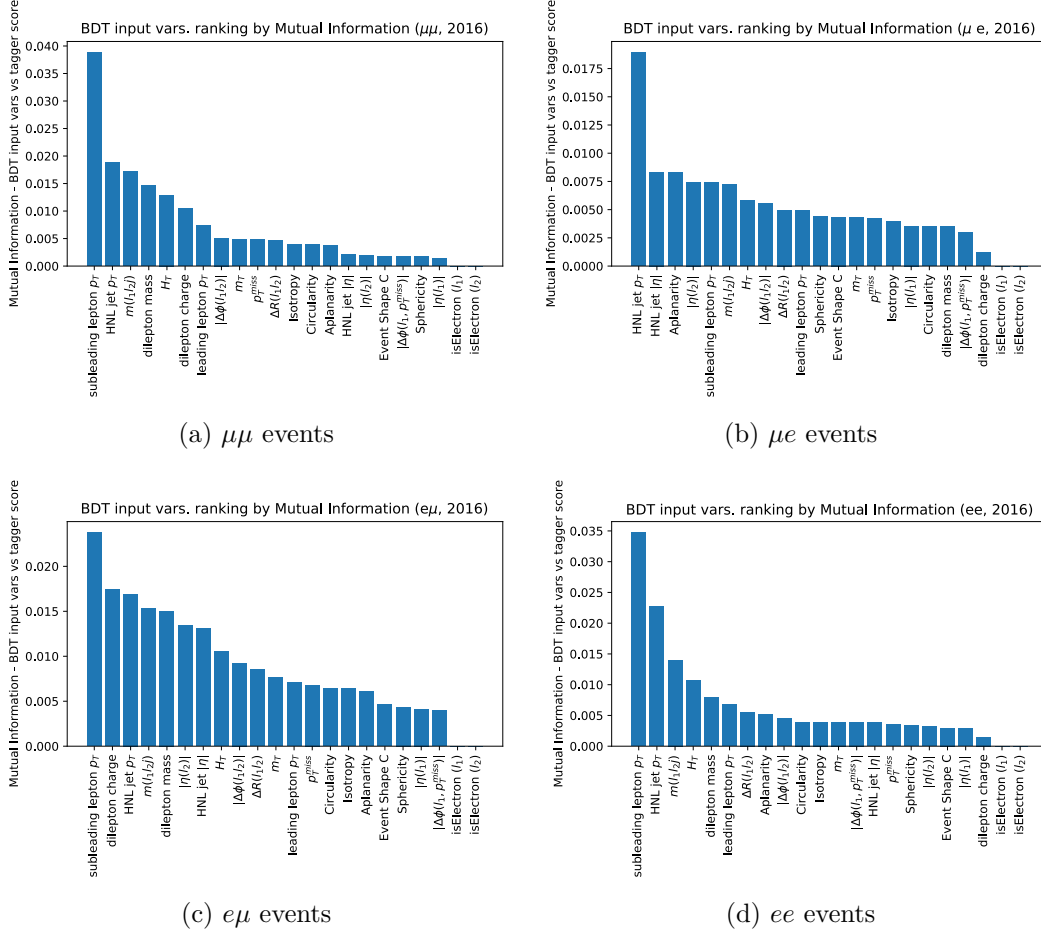


Figure 8.6: BDT input feature ranking the by extent of correlation with the tagger score, using the mutual information metric (2016 simulation).

of MC background events was randomly selected such that the overall class balance between signal and background processes was 50-50% during training, in order to avoid class imbalance. Within the background class, individual MC processes were subsampled based on their expected relative contribution to the signal region counts.

The overall yield for all background processes in the signal region that passed preselection cuts, inclusive of the dilepton flavour and charge categories, was evaluated. The relative ratios between the individual background processes within the SR were then calculated for each data-taking year and plotted in a pie chart (Fig. 8.7). To keep the training and evaluation sets statistically independent, the background events were sampled from MC data sets with LO precision in the accompanying preshower processes: W +jets, DY + jets and $t\bar{t}$ /single- t . The various background processes were sampled according to the fractions listed in Table 8.2. Besides having access to a statistically independent simulation set, using solely LO samples for the training phase has the advantage that no events have negative generator weights, and therefore no further reweighting was needed. Processes below 5% occurrence

rate were removed from the BDT training sample, and the fractions of the remaining background processes were recalculated.

Table 8.2: Predicted relative ratios between background MC processes per data-taking year used in order to subsample the background simulation sets for the BDT training.

Process			
year	2016	2017	2018
$\bar{W} + \text{jets}$	63.5%	65.1%	64.5%
$DY + \text{jets}$	30.3%	29.2%	30.0%
$t\bar{t}/\text{single-}t$	6.2%	5.7%	5.5%

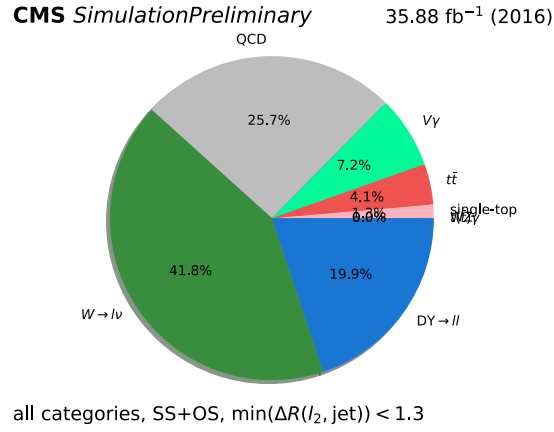
It can be seen that QCD multijet events are a substantial background process of $\sim 25\%$ of all background processes. However, the simulated QCD multijet samples available for MVA training did not contain enough events after passing all the dilepton preselection cuts. Consequently, QCD multijet samples were omitted from the BDT training process. Nevertheless, the BDT classifier was also evaluated using the QCD multijet samples available for the analysis. Since the *uncorrelated BDT* classifier was used along with the displaced jet tagger, their performance was expected to be orthogonal and, therefore, the number of QCD events should be reduced by the tagger.

8.4 The boosted decision tree model evaluation and validation

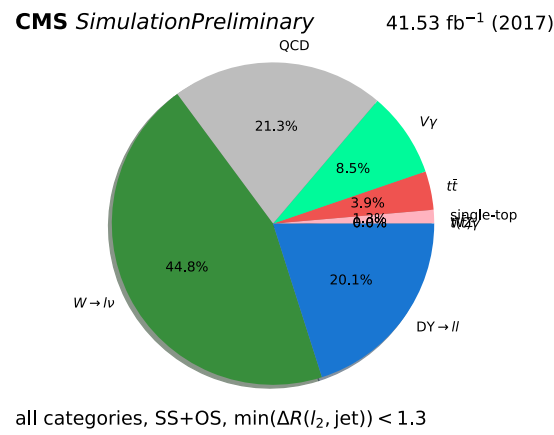
To train the BDT event classifier, the `XGBClassifier` [130] library was used with the logistic loss function introduced in Sec. 5.2.1, a learning rate of 0.05 and 1000 estimators (weak learners). Early stopping was applied to avoid overtraining the algorithm. The simulation set is separated into training (80%) and test samples (20%).

The training and test set classification error and logarithmic loss are monitored, observing them decrease as more weak learners are added during the training. Early stopping is applied if the test sample loss stops decreasing for 10 epochs in order to avoid overtraining (Figs. 8.8-8.9). A similar set of figures was developed for the *correlated BDT* model, presented in Appendix F.4.

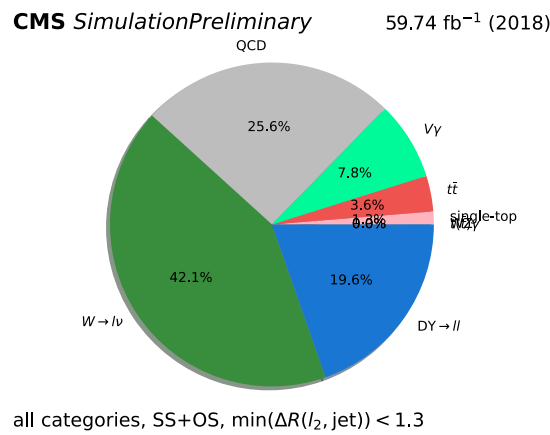
A feature selection study was conducted to train a BDT classifier independent of the tagger and $m(\ell_1, \ell_2, j^*)$ variable (*uncorrelated BDT*), and to find the best-performing classifier, despite its correlation with the tagger information (*correlated*



(a)



(b)



(c)

Figure 8.7: The relative background ratios of MC simulated events in the signal region, for each data-taking year. All dilepton flavour and charge categories are combined for each year. For the BDT training, simulated background processes with LO precision were employed.

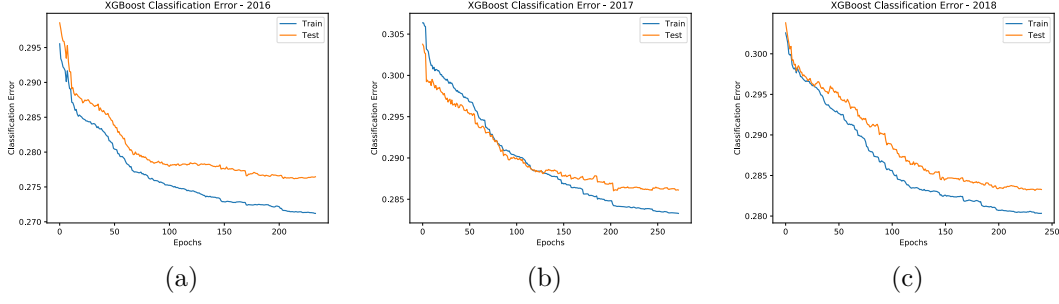


Figure 8.8: Classification error development by training epoch for the uncorrelated BDT model, as well as in the test data set.

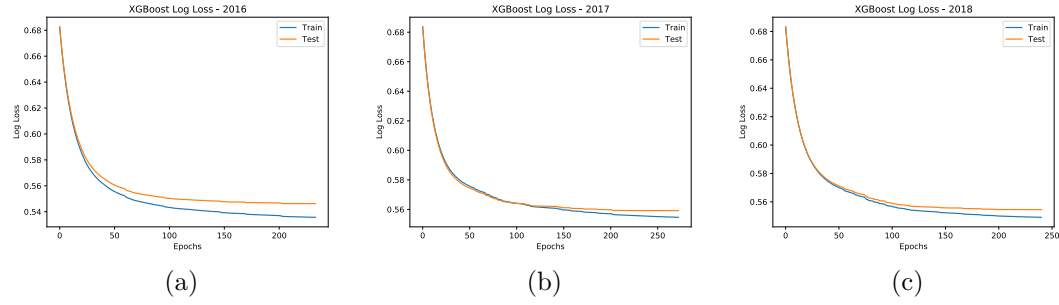


Figure 8.9: Logarithmic loss development by training epoch for the uncorrelated BDT model, as well as in the test data set.

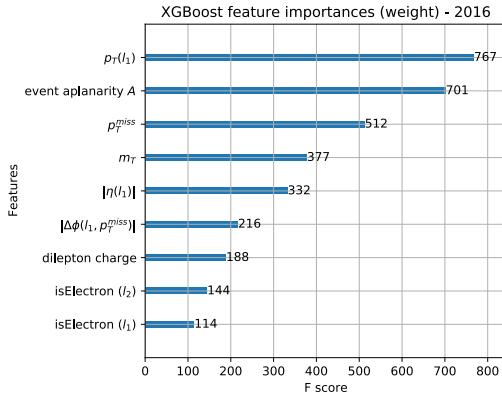
BDT). The input features for the two alternative BDT classifiers are listed in Table 8.3. Event variables depending on the HNL jet candidate kinematics, dilepton and subleading lepton properties are removed from the uncorrelated BDT model, minimising its dependence on the tagger score. The relative importance of the BDT input features during training is shown in Fig. 8.10, measured by the *weight* metric, and showing a different ranking for the two different BDT models. Some of the most discriminating variables in the correlated BDT model are the $m(\ell_1, \ell_2, j^*)$, $p_T(\ell_2)$, H_T , $m(\ell_1, \ell_2)$, $p_T(j^*)$, $|\Delta\phi(\ell_1, \ell_2)|$ and $\Delta R(\ell_1, \ell_2)$, strongly depending on the j^* kinematics and dilepton features of the event. Removing these crucial variables degrades the performance of the BDT but also removes any correlation with the displaced jet tagger score.

Fig. 8.11 shows the strength of the correlation between the DNN tagger output and the uncorrelated BDT score illustrated via 2D heatmaps, giving confidence in the relatively low correlation between the two discriminating variables used in the ABCD data-driven background estimation technique.

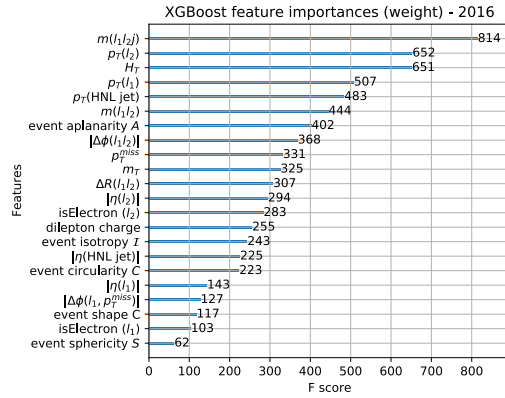
Three HNL signal models with largely different topologies were selected to test the performance of the BDT. The masses of the three HNL benchmark models are 4.5, 8

Table 8.3: Input variables used in the two BDT models considered. The first BDT model considered is uncorrelated with the displaced jet tagger and $m(\ell_1, \ell_2, j^*)$ variable, whereas the second BDT model is highly dependent on them.

Input features	Uncorrelated BDT	Correlated BDT
$m(\ell_1, \ell_2, j^*)$		✓
$m(\ell_1, \ell_2)$		✓
$\Delta\phi(\ell_1, \ell_2)$		✓
$\Delta R(\ell_1, \ell_2)$		✓
dilepton charge	✓	✓
$ \Delta\phi(\ell_1, \vec{p}_T^{\text{miss}}) $	✓	✓
$m_T(\ell_1, \vec{p}_T^{\text{miss}})$	✓	✓
H_T		✓
p_T^{miss}	✓	✓
$p_T^{\ell_1}$	✓	✓
$p_T^{\ell_2}$		✓
$p_T^{\text{HNL jet}}$		✓
$ \eta^{\ell_1} $	✓	✓
$ \eta^{\ell_2} $		✓
$ \eta^{\text{HNL jet}} $		✓
flavour of the leading lepton	✓	✓
flavour of the subleading lepton	✓	✓
isotropy (I)		✓
circularity (C)		✓
sphericity (S)		✓
aplanarity (A)	✓	✓
event shape C		✓



(a)



(b)

Figure 8.10: Feature importances weighted by *weight* for the two BDT models considered (2016 simulation): left plot corresponds to the uncorrelated BDT; right plot corresponds to the correlated BDT.

and 12 GeV, while their lifetimes are 100, 10 and 0.1 mm, respectively. The BDT performance is assessed by plotting the Receiver Operating Characteristic (ROC)

curve for an independent, high-yield dataset containing the three HNL signal samples separately, as well as for simulated background processes at NLO precision in the SR, intended for evaluation and analysis purposes (see Sec. 7.2). The independent, high-yield HNL samples were produced with a generator filter that required at least one lepton with $p_T > 20$ GeV be present in each event.

The ROC curves in Fig. 8.12 present the background efficiency against signal efficiency for the uncorrelated BDT model output for the three independent HNL benchmark models separately for each data-taking year. The area-under-the-curve (AUC) is the quantitative method to assess the performance of the classifier model. In this case, the AUC is defined as the area between the ROC curve and the y-axis. Hence, the BDT optimisation aims to reduce the AUC value; the higher the signal efficiency for a given background efficiency, the better the discriminating power the BDT model has. The uncorrelated BDT model's performance showed an AUC of around 0.17 for the three signal samples and the three separate years. The step-shapes observed in the ROC curves are due to the statistically limited QCD multijet samples with relatively large event weights. The ROC curves for the correlated BDT model are presented in Appendix F.4, showing slightly better performance.

To compare the BDT performance against a simple "cut-and-count" approach, a variable scan study was performed plotting the ROC curves for all BDT input variables separately for each of the three HNL signal samples. Fig. 8.13 shows that the most performant individual variables are p_T^{miss} , isotropy and $m_T(\ell_1, \vec{p}_T^{\text{miss}})$.

The ROC curve corresponding to the p_T^{miss} variable scan was plotted in the same Figure as the BDT ROC curves for all three signal samples considered. A curve corresponding to equal signal and background efficiency equivalent to random guess is also shown in the same Figure. The ROC curves resulting from logistic regression, one of the simplest methods of MVA classification algorithms, were overlaid in the same Figure, as well. The BDT outperforms all the other classifiers, as expected. Furthermore, a baseline cut-based approach was applied for comparison with the following criteria: $H_T < 180$ GeV; $m_T(\ell_1, \vec{p}_T^{\text{miss}}) < 80$ GeV and $p_T^{\text{miss}} < 80$ GeV giving a signal efficiency of 90-95% at a background efficiency of about 70% for all three HNL signal samples. The signal and background efficiency at this working point gave a reduced performance compared to the BDT classifier but better performance compared to the simple p_T^{miss} scan. The background efficiency at a BDT working point of 0.7 was about 15%, whereas the signal efficiency was about 45% for the uncorrelated BDT model.

The BDT model scores were evaluated for both signal and background events in the independent evaluation data set and presented in Fig. 8.14 separately for each

dilepton flavour, charge and separation (boosted vs resolved) category. A single HNL signal model was used for this study with $m_{HNL} = 8$ GeV, $c\tau_0 = 10$ mm, whereas background processes were weighted by their event weight. To avoid fluctuations in the distributions from low-yield, high event weight QCD events, the weights for this process were set to 1 in this study. Good separation between signal and background was observed for most categories, with the exception of opposite-sign same-flavour resolved dilepton categories, where the DY+jets sample is the most dominant SM background process, which the uncorrelated BDT struggled to separate signal events from. A slightly better performance was observed in the case of SS categories as compared to the OS categories.

Finally, the BDT model scores were evaluated in the analysis CRs for MC and data. Figs. 8.15-8.16 present the validation of the BDT output in the two CRs for four representative dilepton event categories. Good data-MC agreement is observed overall, with only minor fluctuations outside statistical and systematic uncertainties. Potential sources for the mild mismodelling issues at higher BDT scores are the low statistics in the QCD multijet simulated samples, as well as the BDT struggling to separate DY+jets background events from the signal in the OSSF categories. Nevertheless, such mismodelling issues can be overcome through the use of data-driven background estimation techniques as discussed in Chapter 10.

8.5 Summary

This Chapter detailed the development of an event-level binary classifier employing a BDT algorithm. The BDT event classifier is used in the data-driven background estimation of the analysis, along with the displaced jet tagger output (see Sec. 10.1). The ABCD method requires that the two discriminating variables used for the method need to be statistically independent. Consequently, the BDT classifier was designed to be uncorrelated with the displaced jet tagger, by picking input variables that are independent of it, as well as validating this assumption via 2D heatmaps and a quantitative measure such as the mutual information. The performance of the BDT event classifier was validated in the SR with an independent simulated data set, showing a background efficiency of about 15% at a BDT working point of 0.7, and a signal efficiency of about 45%. Compared to logistic regression or simple cut-based models, the BDT event classifier has the best discrimination power between signal and background processes. Finally, the BDT output was validated in the analysis' CRs, showing good agreement between data and simulation within statistical and systematic uncertainties, providing confidence in the robustness of the method.

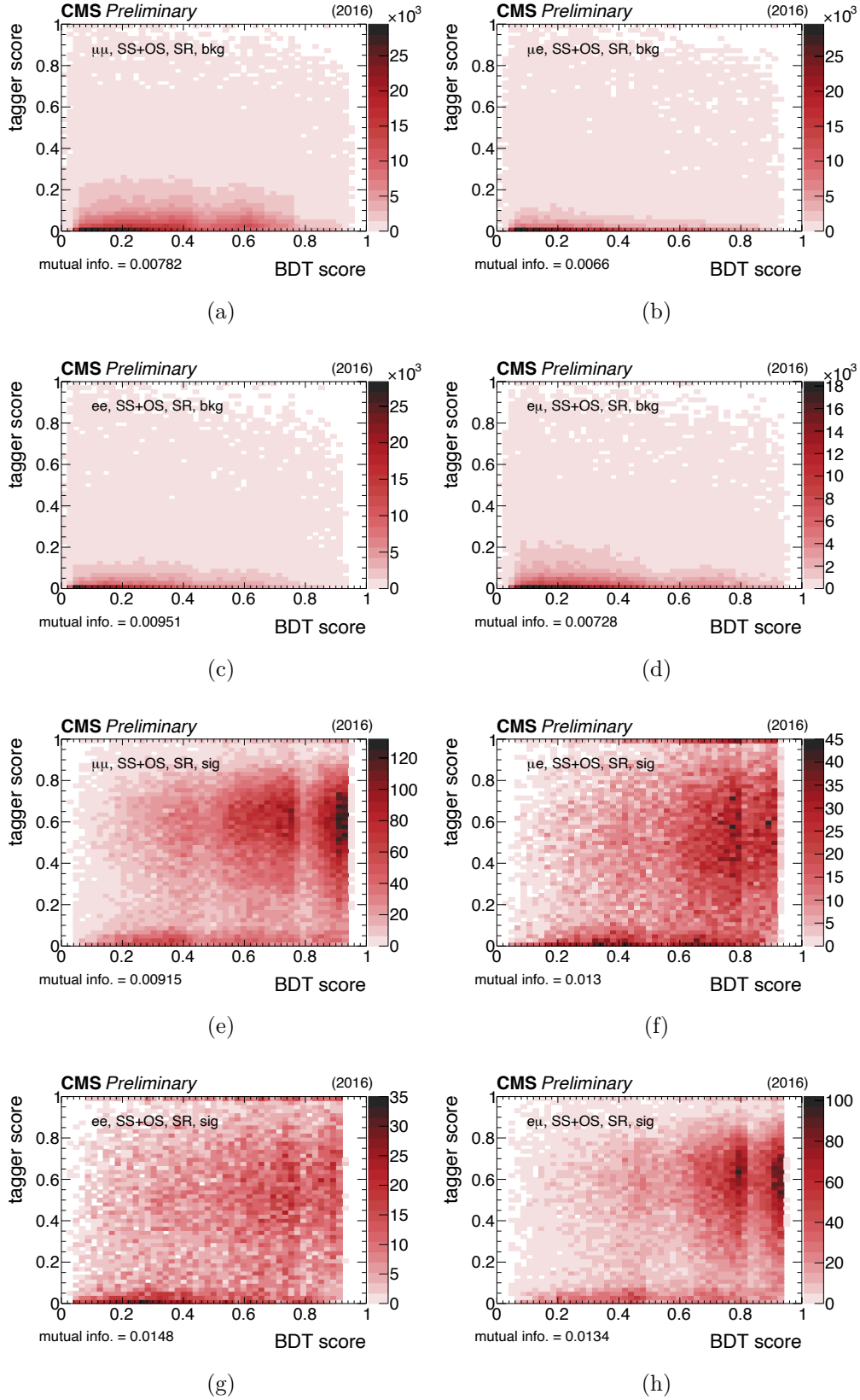


Figure 8.11: 2D heatmaps showing almost no correlation between the score of the uncorrelated BDT model and displaced jet tagger score (2016 simulated background (top four) and signal (bottom four) processes), categorized by dilepton flavour and combined dilepton charge in the SR of the HNL search.

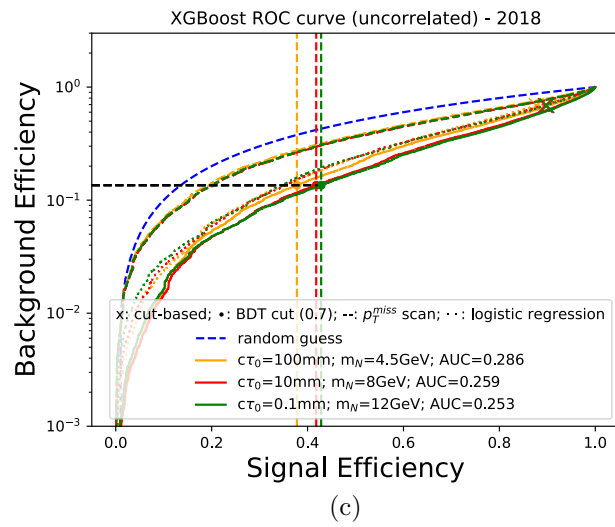
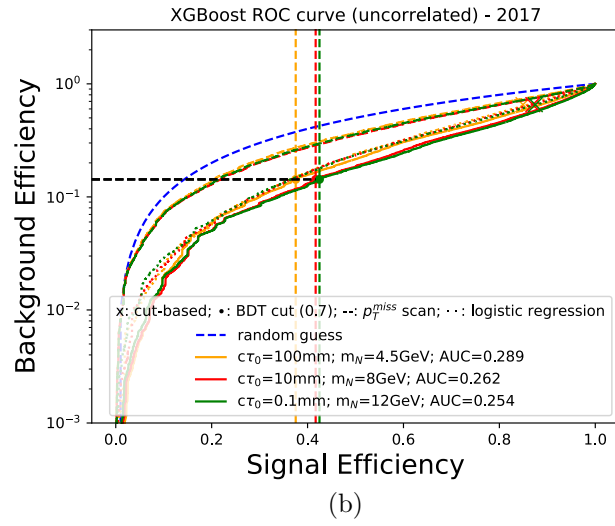
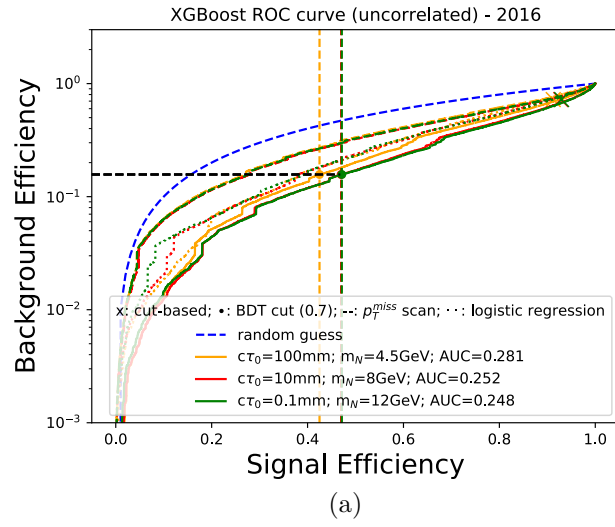


Figure 8.12: BDT performance ROC curves on the evaluation data set for the uncorrelated BDT, presented per separate data-taking years.

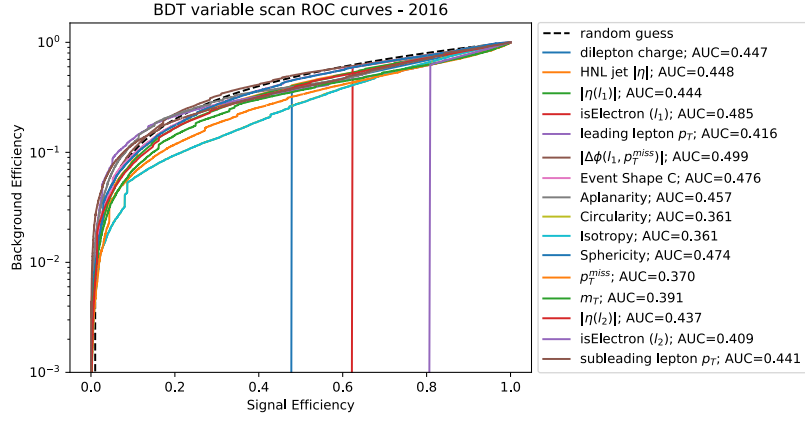
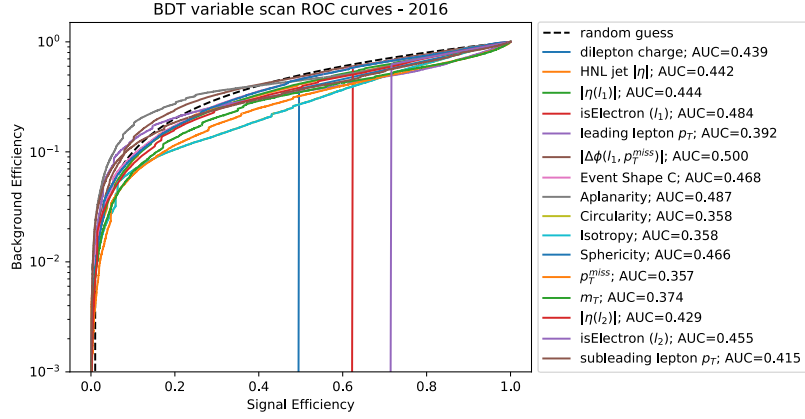
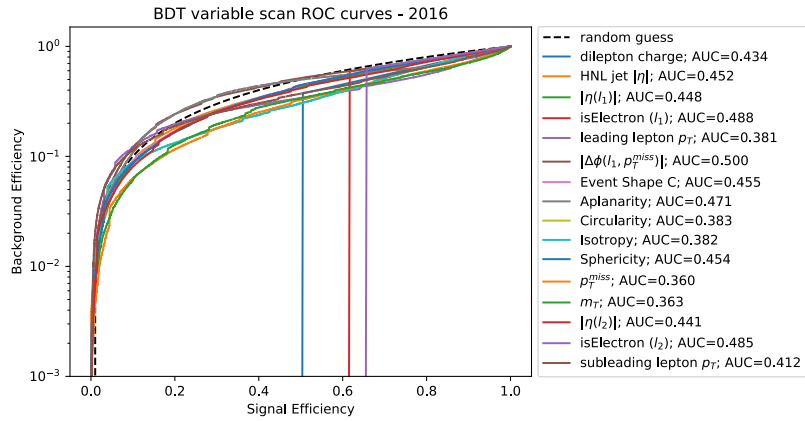
(a) $m_{HNL} = 4.5$ GeV, $c\tau_0 = 100$ mm(b) $m_{HNL} = 8$ GeV, $c\tau_0 = 10$ mm(c) $m_{HNL} = 12$ GeV, $c\tau_0 = 0.1$ mm

Figure 8.13: Variable scan performed on all BDT input features for all three HNL benchmark samples used for evaluation studies, employing the 2016 data set.

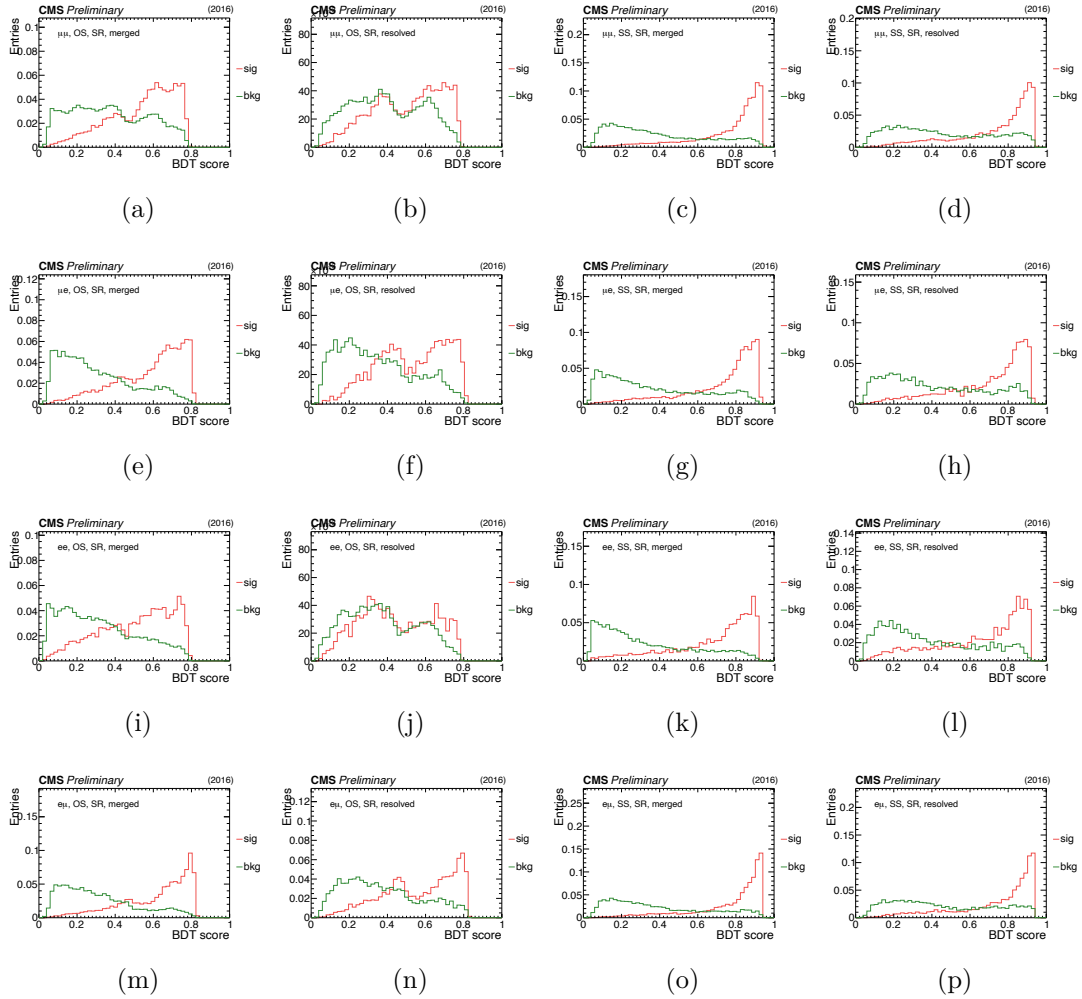


Figure 8.14: BDT score distributions for the uncorrelated model, showing good separation between background and signal processes in most categories (based on 2016 MC simulation).

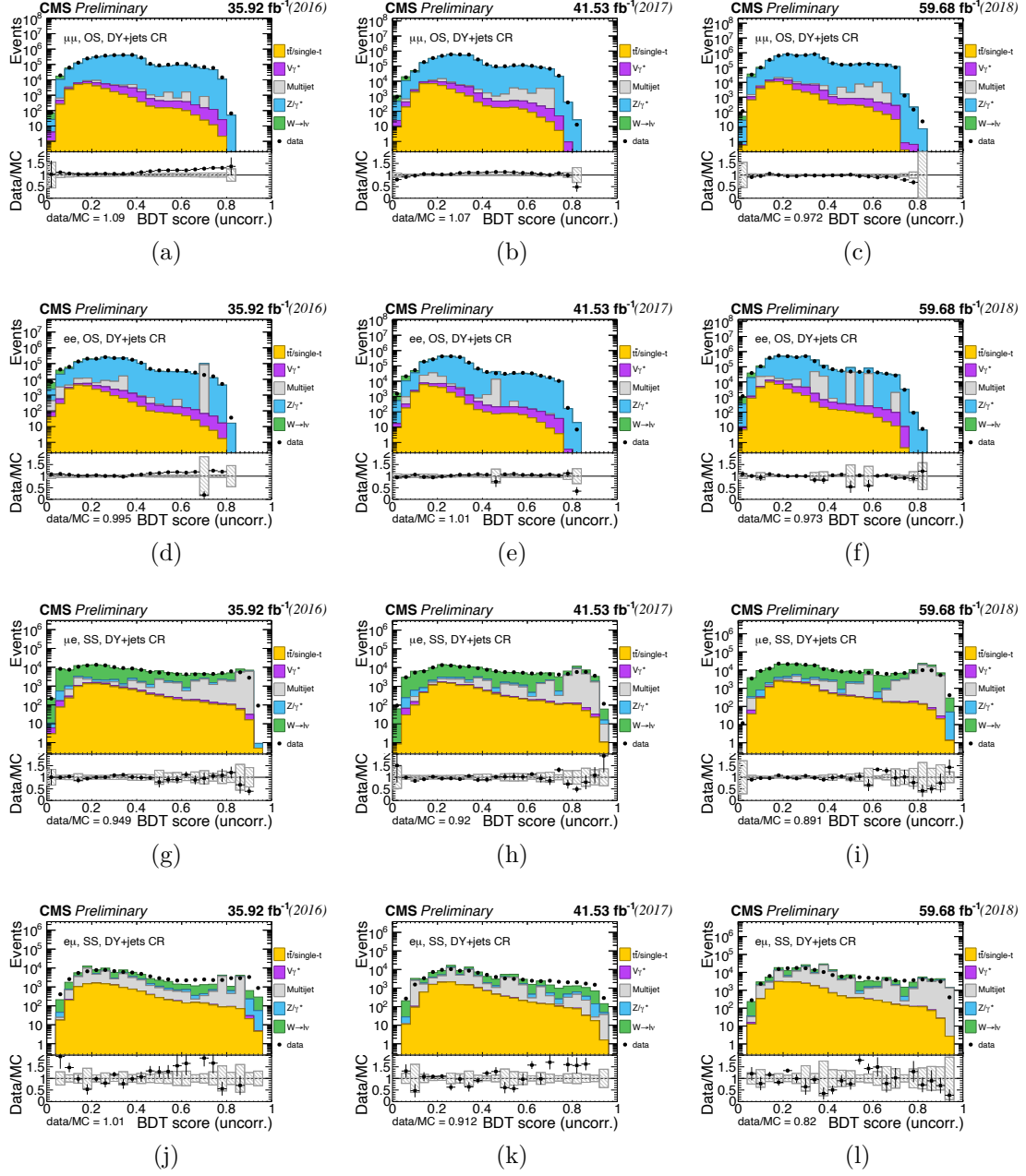


Figure 8.15: Distribution of the uncorrelated BDT model score in the DY control region. The distributions are shown for the 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

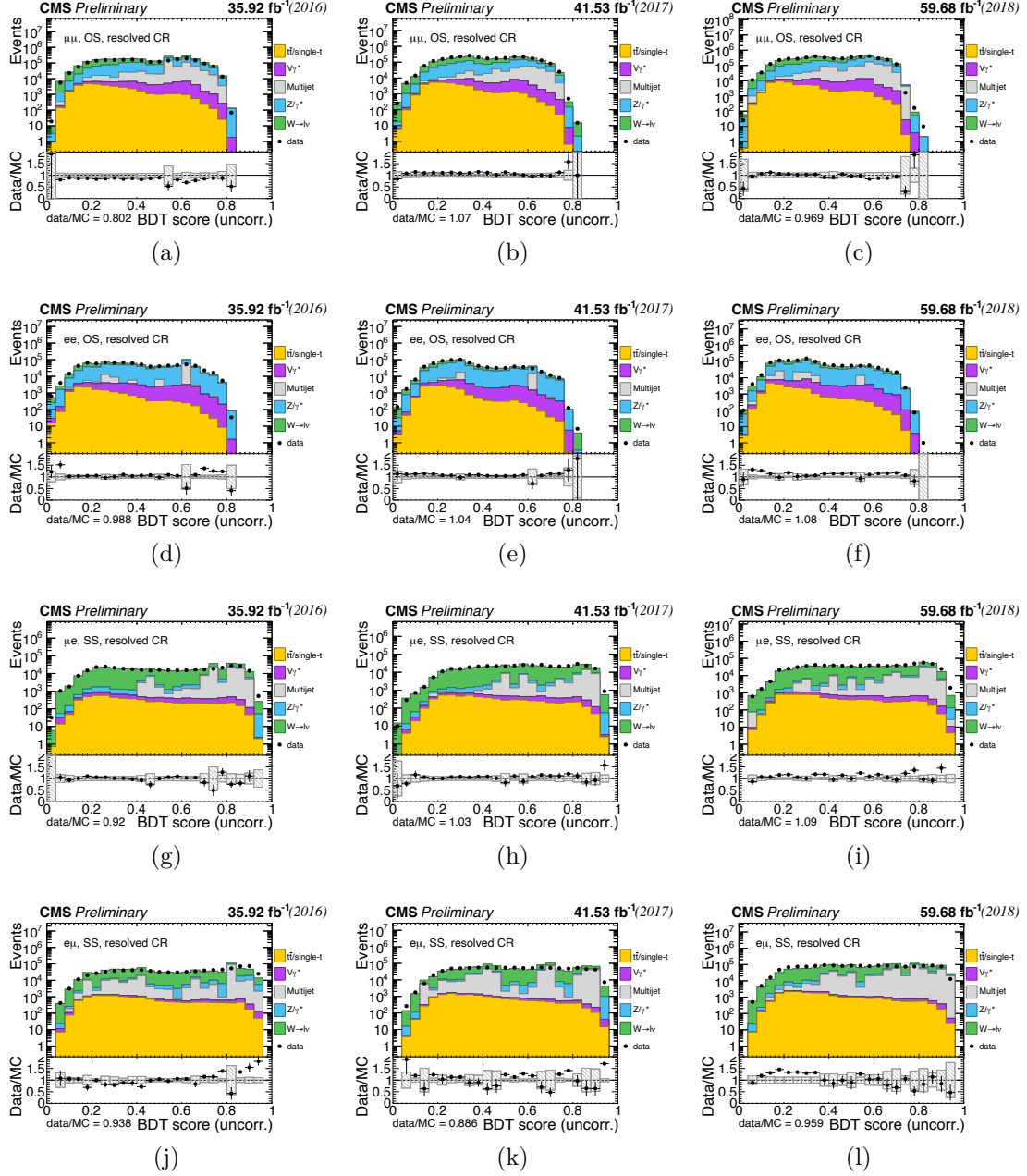


Figure 8.16: Distribution of the uncorrelated BDT model score in the resolved control region. The distributions are shown for the 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

Chapter 9

Displaced lepton systematic uncertainty studies for the long-lived heavy neutral lepton search

This Chapter presents the main sources of systematic uncertainty considered in the search for long-lived HNL particles. Most systematic uncertainties discussed in Sec. 9.1 have an impact on the predicted HNL signal yield in the blinded region of the analysis, whereas the expected background yield is estimated directly from data. Sec. 9.2 discusses the methods used in estimating the efficiencies and scale factors applied to displaced muon identification and reconstruction performance of the CMS detector. In addition, displaced electron identification efficiencies and scale factors are determined in Sec. 9.3. The results of these studies are implemented as corrections to the HNL signal models.

9.1 Sources of systematic uncertainty in the search for long-lived heavy neutral leptons

This Section presents the sources of systematic uncertainty applied in the HNL analysis as nuisance parameters in the statistical evaluation. The majority of the considered systematics impact the signal hypothesis, most of them being derived from simulation, whereas the expected background yields are evaluated via the ABCD method, independently of the systematic uncertainties listed below. The

impacts of these nuisance parameters are profiled for background-only, as well as background+signal hypotheses via maximum likelihood fits to data. Uncertainties were treated to be uncorrelated between the different data-taking years. The size of the effect on the HNL signal cross-section of each listed systematic uncertainty is quantified in Table 10.1 via percentage impacts.

9.1.1 Signal systematic uncertainties

- **Prompt lepton efficiencies** Systematic uncertainties on the reconstruction, identification and trigger efficiency of the prompt lepton candidates are taken into account by varying the applied p_T and η -dependent scale factors within their uncertainties.
- **Displaced lepton efficiencies** Systematic uncertainties on the identification efficiency of the displaced lepton candidates are taken into account by varying the applied p_T and $|d_{xy}^{sig}|$ -dependent scale factors within their uncertainties. Their determination is given in Sections 9.2.1 and 9.3. In addition, systematic uncertainties on displaced muon reconstruction are considered by varying the p_T and $|d_{xy}|$ -dependent scale factors within their uncertainties, which are derived in Section 9.2.2. Systematic uncertainties on the displaced electron reconstruction efficiencies are accounted for by varying the p_T and η -dependent scale factors within their uncertainties.
- **Pileup** A systematic of 5% is applied to the expected pileup contribution in data, derived from minimum bias cross-section.
- **Jet energy scale (JES)** Systematic uncertainties on the jet energy scale in simulated jets are taken into account by varying the applied p_T and η -dependent scale factors within their uncertainties. The resulting discrepancies are propagated to the missing transverse momentum.
- **Jet energy resolution (JER)** Systematic uncertainties on the jet energy resolution are derived by varying the difference in the p_T distribution between the simulated jets and their corresponding generator-level counterparts. If no generator-level jets are matched to the reconstructed one within $\Delta R < 0.4$, a random smearing is applied.
- **Unclustered energy** Missing transverse energy that is unaccounted for by the JES and JER corrections is varied within its uncertainty.

- **Parton distribution function** Systematic uncertainties on the parton distribution functions (PDF) are derived by the prescribed variations on PDF set [126].
- **Factorisation and renormalisation scale** The simulation samples are adjusted by a precomputed set of weights from the event generators, varying the factorisation and renormalisation scales by factors of 2 or 0.5. The overall impact on normalisation is approximately $\pm 12\%$ and is treated as part of the theoretical uncertainty. Factorisation scale factors are used in perturbative quantum field theory to separate long-distance and short-distance physics, while renormalisation scale factors are used to regulate and absorb divergences in quantum field theory calculations.
- **Luminosity** The uncertainties on the measured integrated luminosity per data-taking year (2016-2018) are 1.2%, 2.3%, and 2.5%, respectively [163–165].

9.1.2 Background systematic uncertainties

Uncertainties on the background prediction include a statistical component from the data yields in the sideband regions used in the extrapolation and an additional scaling factor per category derived from closure test studies (see Sec. 10.1).

9.2 Displaced muon efficiencies

This Section presents a detailed description of the strategies used in the HNL search to evaluate efficiencies and scale factors applied as corrections to the signal displaced muon in long-lived HNL decays. The main corrections considered for displaced muons are on their identification and reconstruction efficiency in a 2D kinematic phase-space.

9.2.1 Displaced muon identification efficiency

This Section presents the strategy used in the HNL search to evaluate the displaced muon identification (ID) efficiency, using the tag-and-probe method. Simulated events of $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ are employed to obtain displaced muons mimicking the signature of long-lived semi-leptonic HNL decays. The dimuon pair produced via the J/ψ decay are used for the tag-and-probe method.

The tag-and-probe technique is a widely used data-driven method in high energy physics mostly for calibration and efficiency measurements. A well-known, standard candle resonance process is chosen that is compatible with the phase space of physics objects one aims to measure. Since the subleading leptons in long-lived HNL decays are rather soft and displaced, the $J/\psi \rightarrow \mu_{\text{tag}}^{\pm} \mu_{\text{probe}}^{\mp}$ resonance decay to dimuon final state is used as standard candle. However, J/ψ meson decays are prompt, therefore their direct production cannot be used for this study aiming to capture displaced muons. Instead, one exploits their production via B-meson decays, as mentioned above. In the resonance decay, one muon is being triggered on (i.e. *tagged muon*) with tight identification and reconstruction criteria, while the other muon associated with the resonance (i.e. *probe muon*) has no identification or reconstruction constraints. Having a tracker-based analysis, probe muons are first identified as tracker muons. The yield of tracker muons would be the denominator in the efficiency calculations of the tag-and-probe method whereas those that also pass loose identification requirements, would give the numerator.

$$\varepsilon = \frac{N_{\text{probe}}(\text{ID and reco})}{N_{\text{probe}}(\text{reco})}. \quad (9.1)$$

Differences in the identification efficiencies between data and simulation for a given set of physics observables would yield the scale factors required to be applied as corrections in the analysis. Physics observables considered for displaced muon identification scale factors are related to the kinematic observables and displacement of the muons, for instance, transverse momentum, pseudorapidity, impact parameter or impact parameter significance. Final scale factors were determined for the following parameter spaces: p_T vs $|d_{xy}^{\text{sig}}|$ and p_T vs $|d_{xy}|$. Since the analysis uses a categorisation of events in d_{xy}^{sig} , the scale factor set implemented in the analysis will be of the first pair of variables. These scale factors are applied as corrections to signal MC leptons with a statistical uncertainty of 1σ variation, propagated to the final limits estimation. Scale factors are defined as the fraction between the efficiency in data over the efficiency in simulation as

$$\text{SF}(x, y) = \frac{\varepsilon_{\text{data}}(x, y)}{\varepsilon_{\text{MC}}(x, y)}, \quad (9.2)$$

where x and y represent a pair of physics observables as those mentioned above.

One of the main limitations of this approach is the limited lifetime of B-mesons ($c\tau_B \sim 0.1$ mm), resulting in low statistics in high displacement bins that this analysis aims to probe since some of our HNL signal model lifetimes reach orders of magnitude

higher.

Since this analysis uses default *loose muon* identification criteria for displaced muons without any isolation requirement, there is no need for additional isolation efficiency measurements. Loose ID muons consist of muons reconstructed by both the PF algorithm and as tracker- or global muons (see Sec. 7.4).

Simulated signal samples of the B-meson decay process used in the tag-and-probe study were generated with EvtGen at LO. The single muon trigger selection of the tag-and-probe events ensures the presence of a *tag* muon with $p_T > 8$ GeV and does not bias the selection of the associated *probe* muon: HLT_Mu8. This trigger path is pre-scaled, having between 2.4–8.6 pb⁻¹ of overall luminosity, depending on the year of data collection (see Table 9.1).

Table 9.1: Table presenting the corresponding luminosities for each HLT path in data per data-taking year.

	2016	2017	2018
HLT_Mu8_v	3.9 pb ⁻¹	2.4 pb ⁻¹	8.6 pb ⁻¹

Further pre-selection of the events ensures that the dimuon mass is compatible with the J/ψ meson’s mass ($2.9 \text{ GeV} < m_{\mu\mu} < 3.3 \text{ GeV}$) and both muons are present within the tracker acceptance ($|\mu_{\text{tag,probe}}| < 2.4$). Furthermore, the selection ensures that the tagged muon has at least two hits in the muon chambers and $p_T > 8$ GeV. No requirements on consistency with the B-meson mass between the dimuon system and an additional track are required. Therefore, both prompt and displaced J/ψ particles were selected but due to their different phase-space, they populate different kinematic bins. The overall ID efficiency per bin would not be affected if muons from both displaced and promptly produced J/ψ populate them or only displaced ones.

A J/ψ mass peak is reconstructed for dilepton systems retained from data where the probe muons pass or fail loose identification criteria and is fitted with a relativistic Breit-Wigner with asymmetric running width numerically convoluted with Crystal-Ball function, whereas the combinatorial background is fitted by a falling exponential function. The optimisation algorithm of the fitting function is a global Nelder-Mead (simplex) search with final local gradient descent (migrad). The optimization algorithm uses a binned Poisson log-likelihood loss. For further details on the analytic fitting method please consult [166]–[167]. Figures 9.1–9.2 present the fitted peaks in both simulation and data in the first bin of each parameter pair considered for scale factor in the 2016 data-taking year. The difference between the total fit and background fit gives the yields for the efficiency calculations.

Displaced muon identification efficiencies are evaluated for both data and MC in the

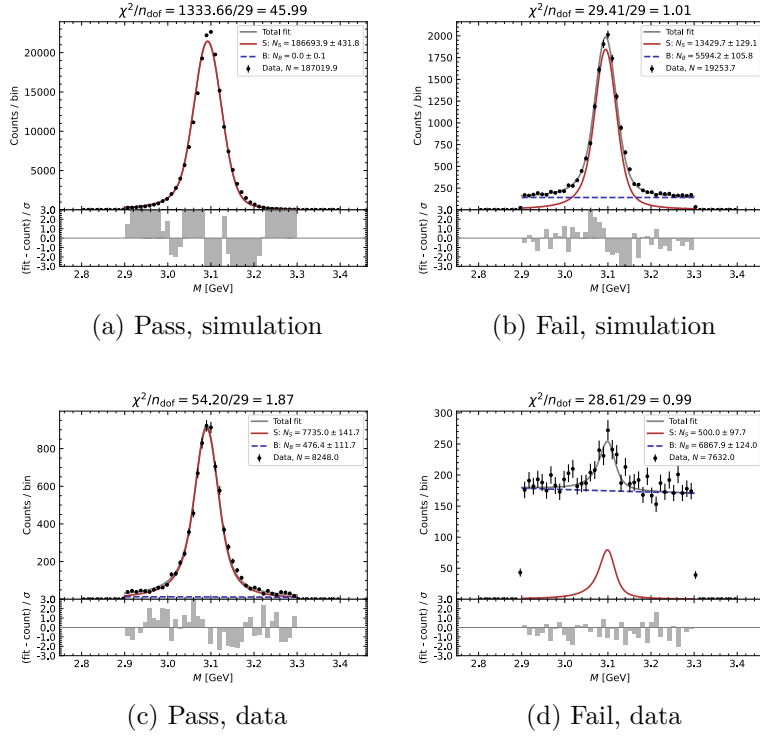


Figure 9.1: Fitted peaks for simulated (top row) and data (bottom row) “pass” and “fail” probe muons in the first bin of the p_T vs $|d_{xy}^{sig}|$ parameter space (2016).

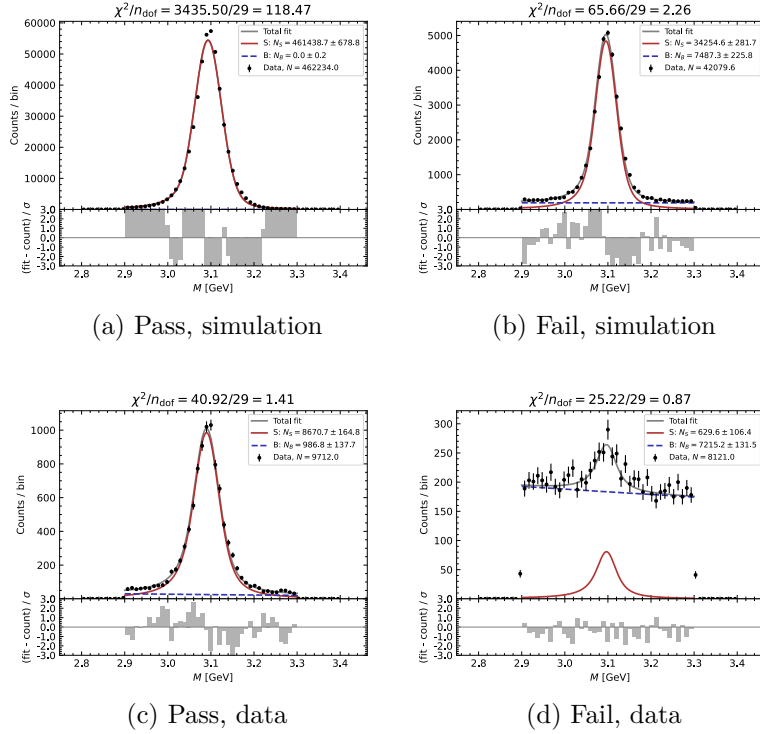


Figure 9.2: Fitted peaks for simulated (top row) and data (bottom row) “pass” and “fail” probe muons in the first bin of the p_T vs $|d_{xy}|$ parameter space (2016).

full 2D phase space of p_T vs $|d_{xy}^{sig}|$ and $|d_{xy}|$ variable distributions (see Figs. 9.3 and 9.4) for all three data-taking years.

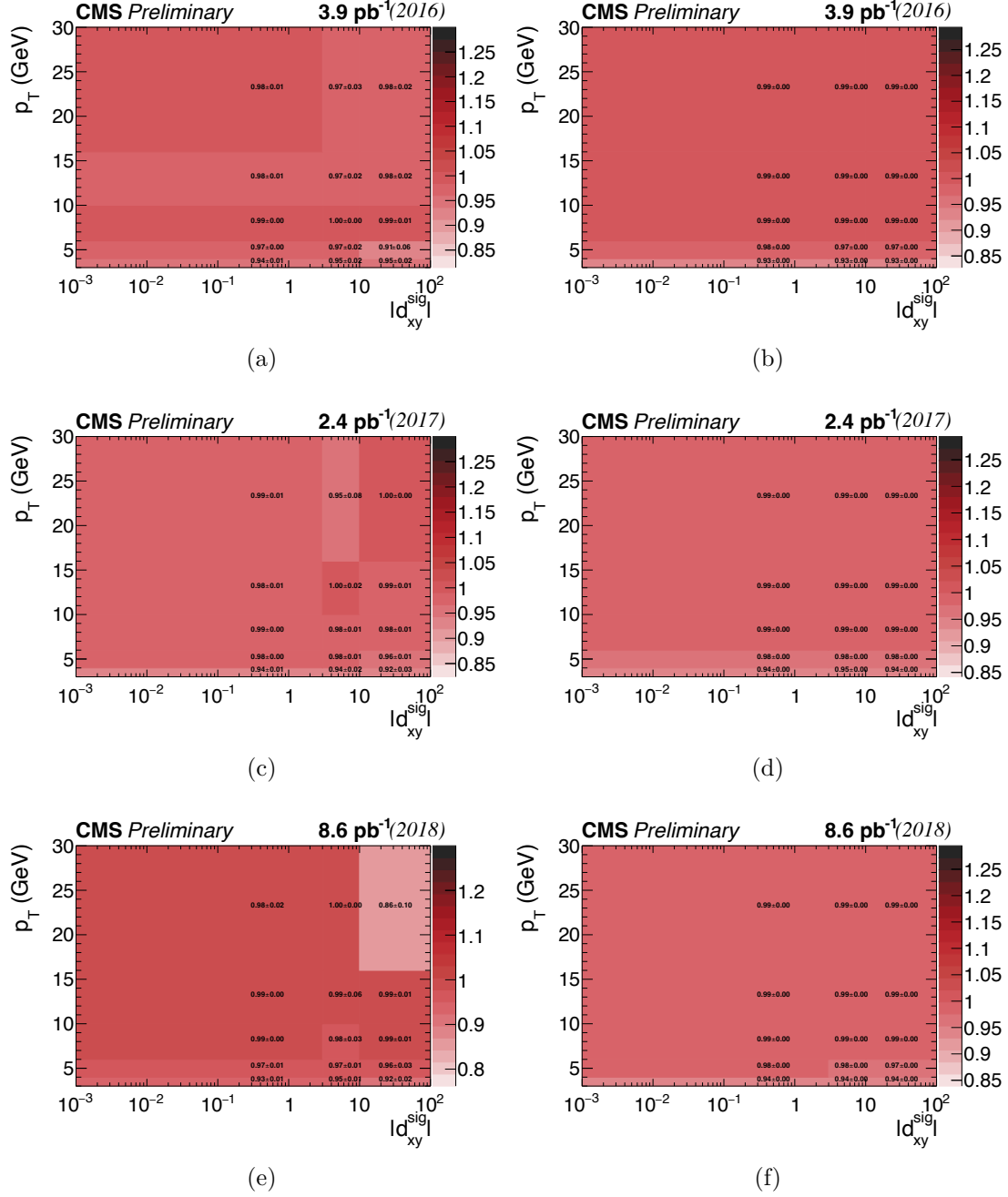


Figure 9.3: Identification efficiencies of data (left) and MC (right) as a function of the probe muon p_T vs $|d_{xy}^{sig}|$ for all 3 data-taking years (2016 to 2018, top to bottom).

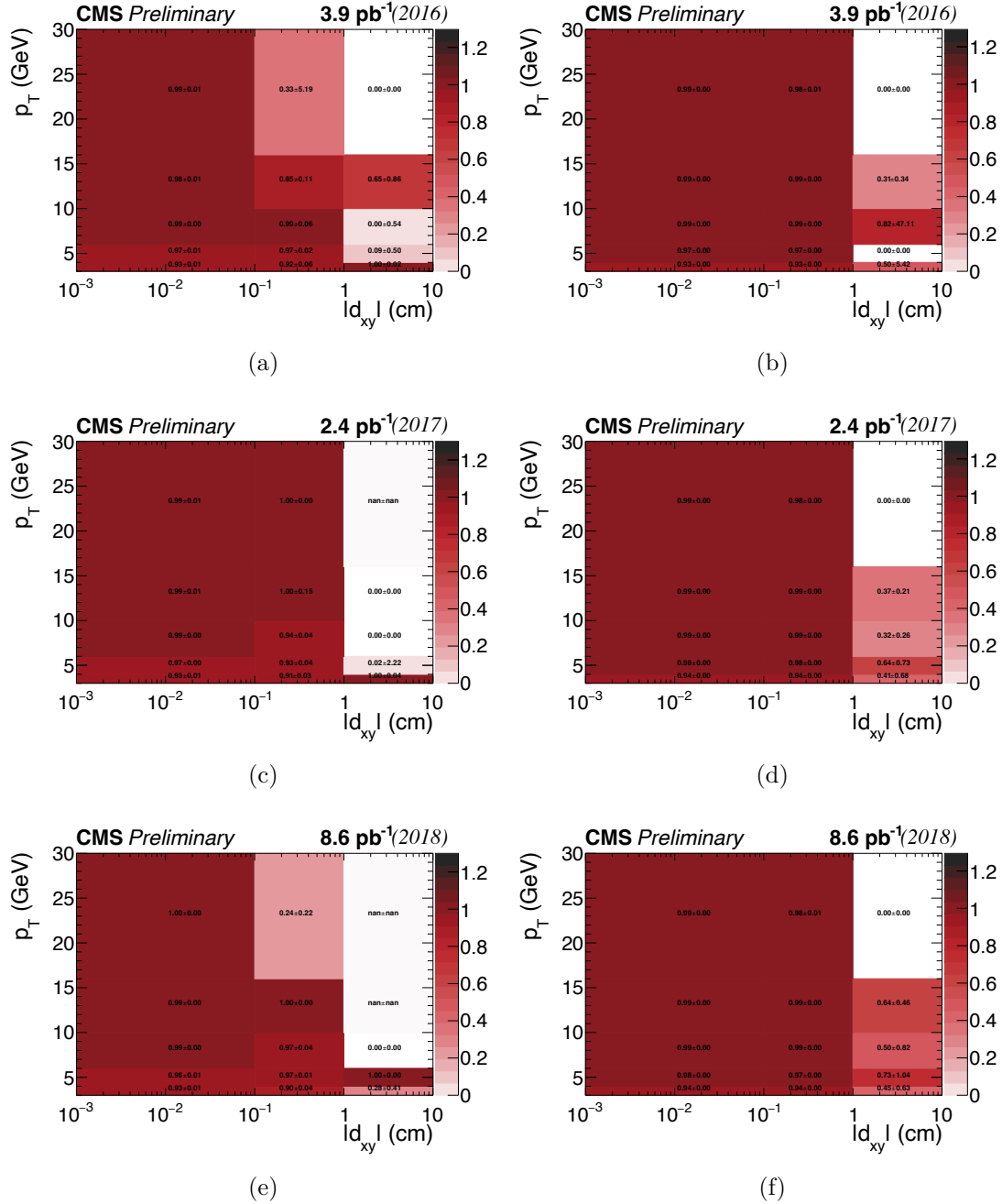


Figure 9.4: Identification efficiencies of data (left) and MC (right) as a function of the probe muon p_T vs $|d_{xy}|$ for all 3 data-taking years (2016 to 2018, top to bottom).

Finally, the scale factors for the displaced muon ID as a function of p_T vs $|d_{xy}^{sig}|$ and $|d_{xy}|$ variable distributions are presented in Tables 9.2 and 9.3 for all three data-taking years (equivalent heatmaps are shown in Appendix D.1). The p_T vs $|d_{xy}^{sig}|$ set of scale factors were implemented as corrections in the analysis framework. The scale factors derived were assigned statistical uncertainties only, using error propagation (see App. A). Most scale factor values are within 2–3% deviations from unity, a promising result regarding the effectiveness of displaced muon identification measurement. Due to the limited statistics in higher displacement bins of B-meson decays, values in the highest $|d_{xy}|$ bin are prone to larger fluctuations from unity with larger associated uncertainty.

Table 9.2: Scale factor values evaluated for displaced muon loose identification criteria as a function of the muon p_T and $|d_{xy}^{sig}|$ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events for all three data-taking years. These values have been implemented in the analysis framework as correction scale factors.

(a) CMS Preliminary, 3.9 pb⁻¹ (2016)

p_T (GeV) vs $ d_{xy}^{sig} $	$10^{-3} - 3.0$	3.0-10.0	10.0-100.0
16.0-30.0	0.99 ± 0.01	0.98 ± 0.03	0.99 ± 0.02
10.0-16.0	0.99 ± 0.01	0.97 ± 0.02	0.99 ± 0.02
6.0-10.0	0.99 ± 0.00	1.01 ± 0.00	1.00 ± 0.01
4.0-6.0	1.00 ± 0.00	0.99 ± 0.02	0.94 ± 0.06
3.0-4.0	1.01 ± 0.01	1.01 ± 0.02	1.02 ± 0.03

(b) CMS Preliminary, 2.4 pb⁻¹ (2017)

p_T (GeV) vs $ d_{xy}^{sig} $	$10^{-3} - 3.0$	3.0-10.0	10.0-100.0
16.0-30.0	1.00 ± 0.01	0.96 ± 0.08	1.01 ± 0.00
10.0-16.0	0.99 ± 0.01	1.00 ± 0.02	0.99 ± 0.01
6.0-10.0	0.99 ± 0.00	0.99 ± 0.01	0.99 ± 0.01
4.0-6.0	1.00 ± 0.00	1.00 ± 0.01	0.99 ± 0.01
3.0-4.0	0.99 ± 0.01	1.00 ± 0.02	0.98 ± 0.03

(c) CMS Preliminary, 8.6 pb⁻¹ (2018)

p_T (GeV) vs $ d_{xy}^{sig} $	$10^{-3} - 3.0$	3.0-10.0	10.0-100.0
16.0-30.0	0.99 ± 0.02	1.01 ± 0.00	0.87 ± 0.10
10.0-16.0	1.00 ± 0.00	1.00 ± 0.06	0.99 ± 0.01
6.0-10.0	1.00 ± 0.00	0.98 ± 0.03	1.00 ± 0.01
4.0-6.0	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.03
3.0-4.0	0.99 ± 0.01	1.01 ± 0.01	0.98 ± 0.02

Table 9.3: Scale factor values evaluated for displaced muon loose identification criteria as a function of the muon p_T and $|d_{xy}|$ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events.

(a) CMS Preliminary, 3.9 pb ⁻¹ (2016)			
p_T (GeV) vs $ d_{xy} $ (cm)	$10^{-3} - 0.1$	0.1-1.0	1.0-10.0
16.0-30.0	1.00±0.01	0.34±5.31	nan±nan
10.0-16.0	0.99±0.01	0.86±0.11	2.08±3.52
6.0-10.0	1.00±0.00	1.00±0.06	0.00±0.68
4.0-6.0	0.99±0.01	1.00±0.02	inf±nan
3.0-4.0	1.00±0.01	0.98±0.06	1.97±21.16
(b) CMS Preliminary, 2.4 pb ⁻¹ (2017)			
p_T (GeV) vs $ d_{xy} $ (cm)	$10^{-3} - 0.1$	0.1-1.0	1.0-10.0
16.0-30.0	0.99±0.01	1.02±0.00	nan±nan
10.0-16.0	0.99±0.01	1.01±0.15	0.00±nan
6.0-10.0	0.99±0.00	0.95±0.04	0.00±nan
4.0-6.0	1.00±0.00	0.96±0.04	0.03±3.45
3.0-4.0	0.99±0.01	0.97±0.03	2.42±3.97
(c) CMS Preliminary, 8.6 pb ⁻¹ (2018)			
p_T (GeV) vs $ d_{xy} $ (cm)	$10^{-3} - 0.1$	0.1-1.0	1.0-10.0
16.0-30.0	1.01±0.00	0.24±0.23	nan±nan
10.0-16.0	1.00±0.00	1.01±0.00	nan±nan
6.0-10.0	1.00±0.00	0.98±0.04	0.00±nan
4.0-6.0	0.99±0.01	1.00±0.01	1.38±1.97
3.0-4.0	0.99±0.01	0.96±0.04	0.61±1.24

9.2.2 Displaced muon track reconstruction efficiency

This Section presents the strategy used in this analysis to evaluate the displaced muon reconstruction efficiency, employing the same tag-and-probe method and software tools as the ones presented in Sec. 9.2.1.

Fig. 4.2 in Sec. 4.2.1 shows a decrease in muon reconstruction efficiency of muons from HNL decays at low generator-matched p_T and increasing transverse displacement, L_{xy} , the latter being correlated with the impact parameter d_{xy} variable used in the rest of the analysis. Consequently, it is expected to have significant deviations from unity on the muon reconstruction scale factors derived in that kinematic phase-space.

The tag-and-probe method used in this study employs the J/ψ dimuon resonance as a standard candle, produced via the same B-meson decay chain as in the muon identification efficiency study presented in the previous section. Both simulation and data samples have been reused for this study, including the same pre-scaled trigger path. The tagged muon which is being triggered on, is defined as a tight, well-reconstructed muon that matches a track. The probe muons in the samples used in this study, as well as the ones in the previous section, are by default muons with matching tracks in the tracker. However, if there is no matching muon for a reconstructed tracker track, the latter is still retained in the sample and the dimuon mass is in fact, the invariant mass of the tagged muon and the probe tracker track. While the previous study only considered tracks that had a matching muon, this study also employs unmatched tracks as probe muon candidates. In the case of evaluating the muon reconstruction efficiencies, the yield of tracker muons would be the numerator in the efficiency calculations whereas the yield of all tracks retained in the samples, either matched or unmatched to muon candidates, would determine the denominator.

$$\varepsilon = \frac{N_{\text{probe}}(\text{tracker muons})}{N_{\text{probe}}(\text{tracks})}. \quad (9.3)$$

This study uses the same event pre-selection, ensuring compatibility with the J/ψ meson's mass peak, as the tag-and-probe method presented in the previous section, with the dimuon mass being required in the $2.9 \text{ GeV} < m_{\mu\mu} < 3.3 \text{ GeV}$ mass range, as well as both the tagged muon and the tracker track being required to be within the tracker acceptance. Due to the HLT path employed, the tagged muon is required to have $p_T > 8 \text{ GeV}$ and at least two hits in the muon chamber.

In addition, the fitting algorithm used in determining the expected pass and fail probe tracks or tracker muons is the same as in Sec. 9.2.1, based on a relativistic

Breit-Wigner function with asymmetric running width numerically convoluted with a Crystal-Ball function, whereas the combinatorial background is fitted by a falling exponential function. Since the impact parameter error and significance variables were not saved for tracker tracks in the samples used in this study, the efficiency calculations and scale factors have been derived on the p_T vs $|d_{xy}|$ variable pair, instead. A set of J/ψ resonance mass fits in a number of bins of interest is presented in Figs. 9.5-9.7. Some fail bins show very little or no signal peak, resulting in unitary efficiencies in such bins, while other bins may be statistically limited.

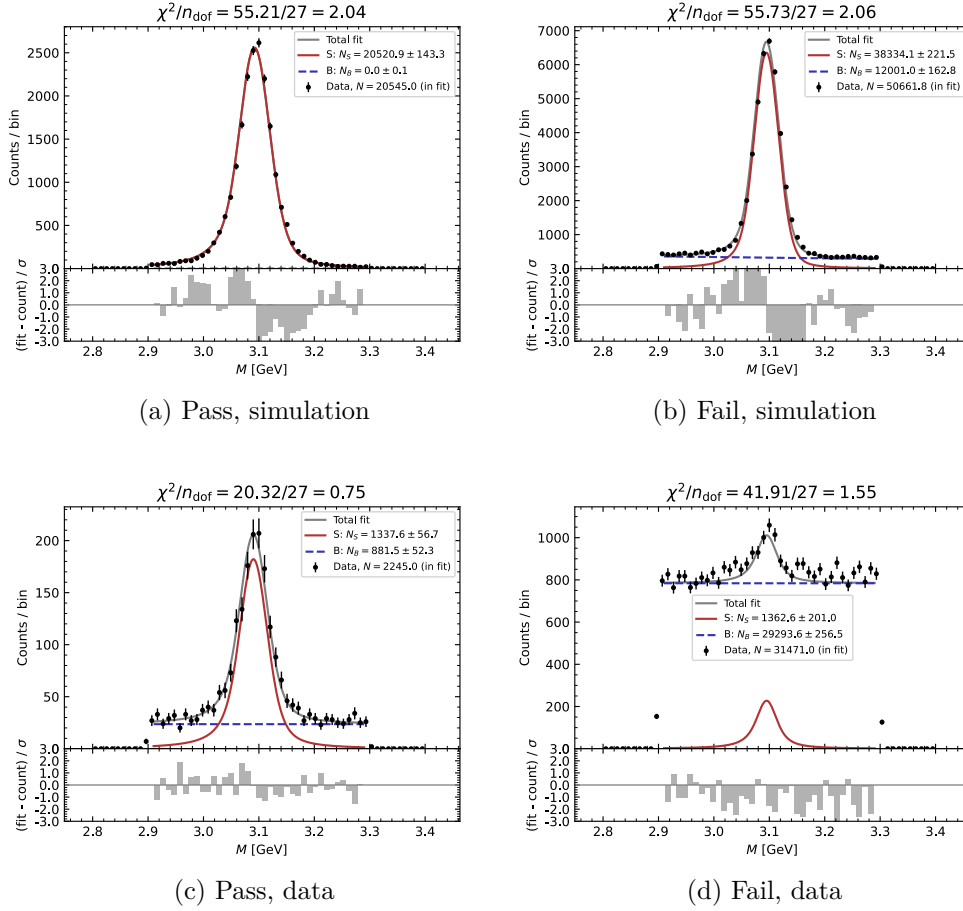


Figure 9.5: Fitted peaks for simulated (top row) and data (bottom row) “pass” and “fail” probe tracker muons and tracks in the first bin of the p_T vs $|d_{xy}|$ parameter space (2016).

Displaced muon reconstruction efficiencies are evaluated for both data and MC in the full 2D phase space of p_T vs $|d_{xy}|$ variable distributions (see Fig. 9.8) for all three data-taking years. As expected, efficiencies show larger deviations from unity in low- p_T bins.

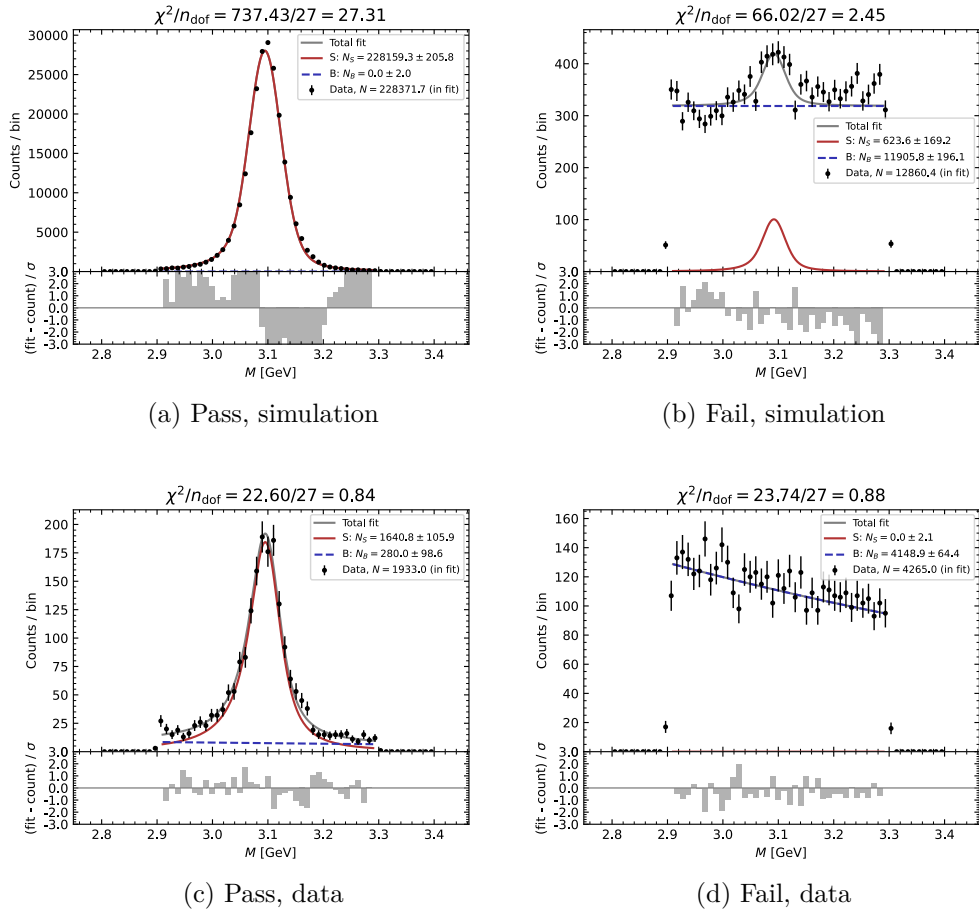


Figure 9.6: Fitted peaks for simulated (top row) and data (bottom row) "pass" and "fail" probe tracker muons and tracks in a mid-range bin of the p_T vs $|d_{xy}|$ parameter space (2016).

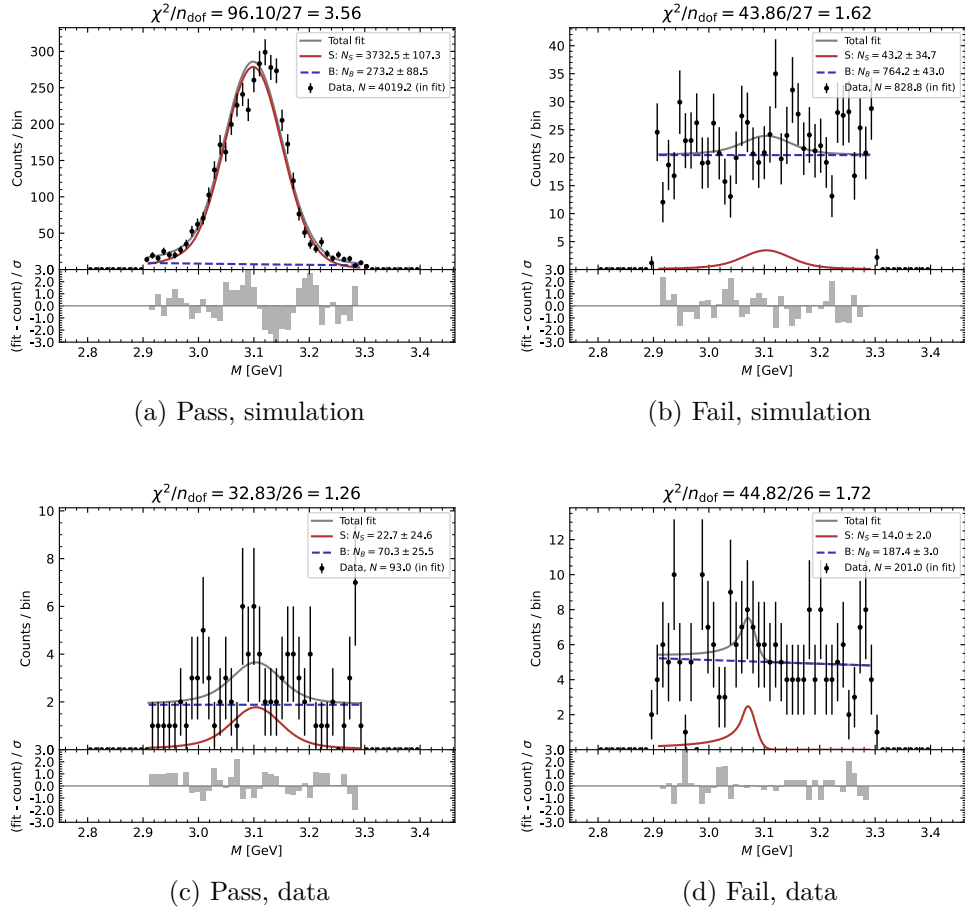


Figure 9.7: Fitted peaks for simulated (top row) and data (bottom row) "pass" and "fail" probe tracker muons and tracks in the highest bin of the p_T vs $|d_{xy}|$ parameter space (2016).

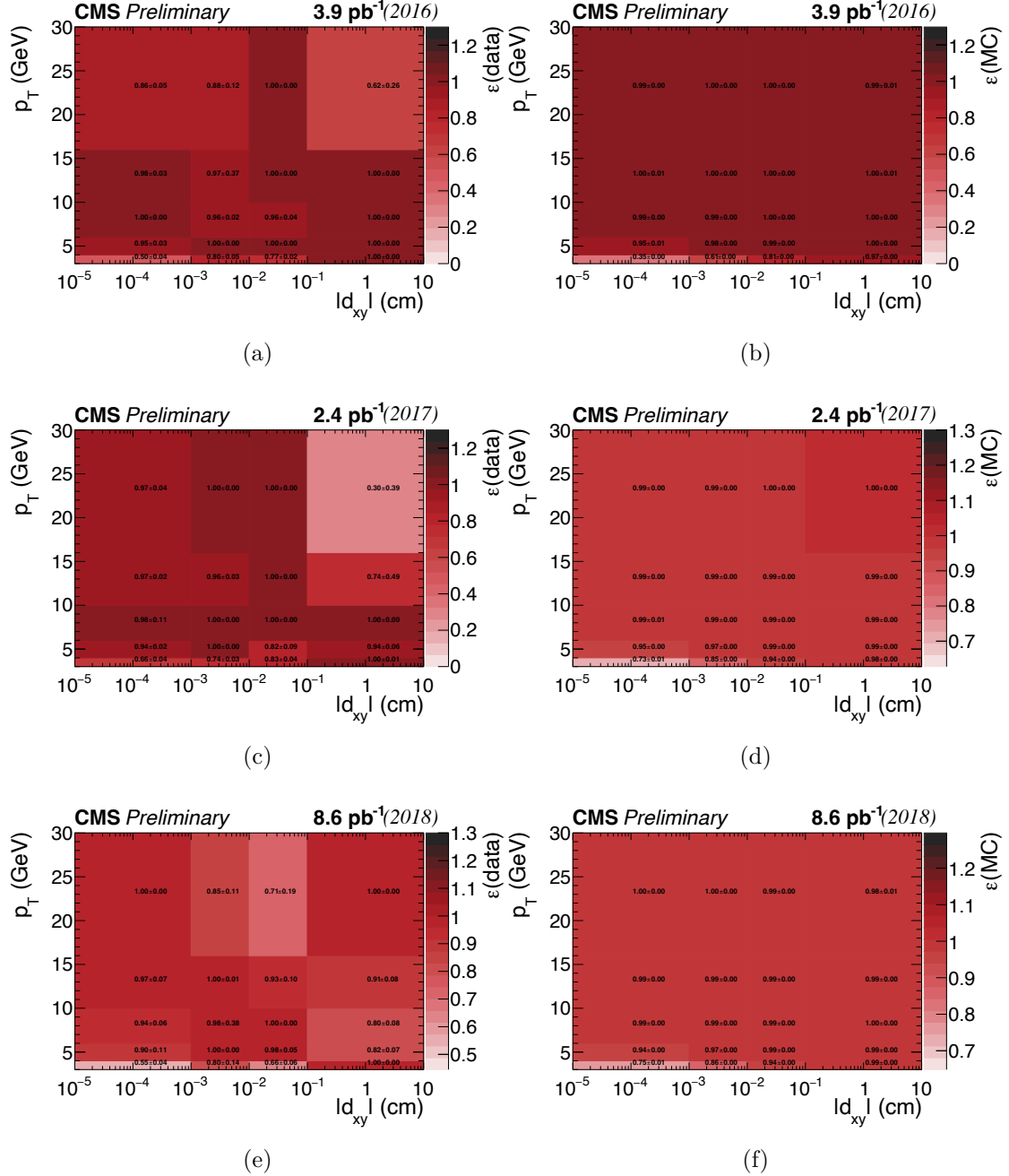


Figure 9.8: Reconstruction efficiencies of data (left) and MC (right) as a function of the probe muon p_T vs $|d_{xy}|$ for all 3 data-taking years (2016 to 2018, top to bottom).

Finally, the scale factors on displaced muon reconstruction as a function of p_T vs $|d_{xy}|$ variable distribution are presented in Table 9.4 for all three data-taking years (equivalent heatmaps are shown in Appendix D.2). These scale factors have been implemented as corrections in the analysis framework, besides the loose muon identification scale factors evaluated in the previous section. The scale factors derived were assigned statistical uncertainties only, evaluated via error propagation. Most scale factor values are within 2–3% deviation from unity, except at low- p_T , as was expected from previous observations in generator-matched studies. Some bins have limited statistics with large resulting statistical uncertainties. Scale factor values of such bins were manually adjusted via extrapolation from nearby bins, for instance, those in the high p_T - $|d_{xy}|$ corner of the 2D kinematic phase-space studied.

Table 9.4: Scale factor values evaluated for displaced muon reconstruction as a function of the muon p_T and $|d_{xy}|$ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events for all three data-taking years. These scale factors are eventually propagated in the analysis framework as corrections.

(a) CMS Preliminary, 3.9 pb ⁻¹ (2016)				
p_T (GeV) vs $ d_{xy} $ (cm)	$10^{-5} - 10^{-3}$	$10^{-3} - 10^{-2}$	$10^{-2} - 0.1$	0.1-10.0
16.0-30.0	0.87±0.05	0.88±0.12	1.00±0.00	1.00±0.00
10.0-16.0	0.98±0.03	0.97±0.37	1.00±0.00	1.00±0.01
6.0-10.0	1.01±0.00	0.97±0.02	0.96±0.04	1.00±0.00
4.0-6.0	1.00±0.03	1.02±0.00	1.01±0.00	1.00±0.00
3.0-4.0	1.42±0.11	1.32±0.08	0.95±0.03	1.03±0.00
(b) CMS Preliminary, 2.4 pb ⁻¹ (2017)				
p_T (GeV) vs $ d_{xy} $ (cm)	$10^{-5} - 10^{-3}$	$10^{-3} - 10^{-2}$	$10^{-2} - 0.1$	0.1-10.0
16.0-30.0	0.98±0.04	1.01±0.00	1.01±0.00	1.01±0.00
10.0-16.0	0.98±0.02	0.97±0.03	1.01±0.00	1.01±0.00
6.0-10.0	0.99±0.11	1.01±0.00	1.01±0.00	1.01±0.00
4.0-6.0	1.00±0.03	1.03±0.00	0.83±0.10	0.95±0.06
3.0-4.0	0.90±0.06	0.87±0.04	0.88±0.04	1.02±0.01
(c) CMS Preliminary, 8.6 pb ⁻¹ (2018)				
p_T (GeV) vs $ d_{xy} $ (cm)	$10^{-5} - 10^{-3}$	$10^{-3} - 10^{-2}$	$10^{-2} - 0.1$	0.1-10.0
16.0-30.0	1.00±0.00	0.98±0.02	0.98±0.02	1.01±0.01
10.0-16.0	0.99±0.08	1.01±0.01	0.94±0.10	0.92±0.08
6.0-10.0	0.95±0.06	0.99±0.38	1.01±0.00	0.81±0.08
4.0-6.0	0.95±0.11	1.03±0.00	0.99±0.05	0.82±0.07
3.0-4.0	0.73±0.06	0.94±0.16	0.70±0.07	1.01±0.00

9.3 Displaced electron efficiencies

In this Section, a detailed description of the strategies used in the HNL search to evaluate efficiencies and scale factors applied as corrections to the signal displaced electrons in long-lived HNL decays is presented. The main corrections considered for displaced electrons are on their identification efficiency in a 2D kinematic phase-space.

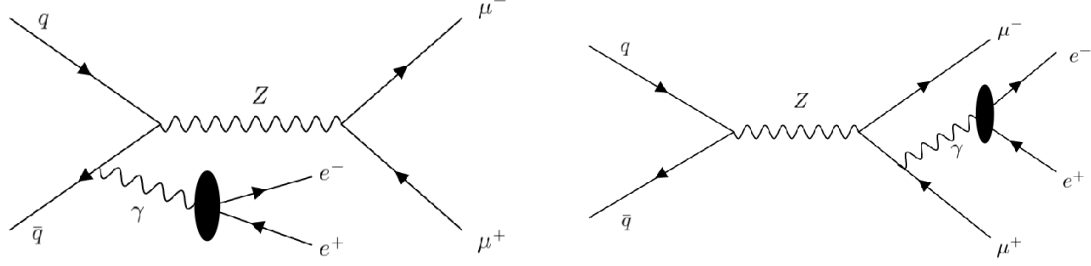
9.3.1 Displaced electron identification scale factors

This section presents the strategy used in the analysis to evaluate the displaced electron identification efficiency, via asymmetric photon conversion in events of

$$\begin{aligned} Z\gamma &\rightarrow \mu^- \mu^+ \gamma \rightarrow \mu^- \mu^+ e^\pm (e^\mp) \\ Z &\rightarrow \mu^- \mu^+ \gamma \rightarrow \mu^- \mu^+ e^\pm (e^\mp), \end{aligned} \quad (9.4)$$

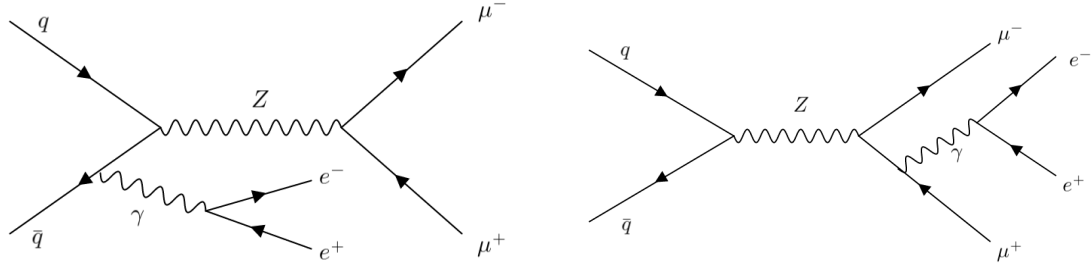
where $(e^\mp)$ represents a low- p_T electron that potentially fails reconstruction and/or identification criteria. In this study, both initial state radiation (ISR) photons produced along with the Z boson ($Z\gamma$) and final state radiation (FSR) photons produced via the Z boson decay ($Z \rightarrow \mu^- \mu^+ \gamma$) are considered (see Fig. 9.9). Such photons can be either external (real) photons that interact with the detector material and hence convert into displaced electron-positron pairs (Figs. 9.9a and 9.9b) that are aimed to be studied, or they may be internal (virtual) photons producing prompt signatures of dielectrons (Figs. 9.9c and 9.9d). Since internal photon conversions into electron pairs occur close to the interaction point at low displacement, they do not serve as intrinsic mismodelling to scale factors capturing higher displacements and hence no special treatment is required. In asymmetric photon conversions, one particle of the electron-positron pair carries a significant fraction of the photon energy, which serves to be a good approximation to the overall photon kinematics, as well as a probe for nonprompt electron efficiency measurements in the detector's tracking material.

This study employs the existing HNL analysis framework for prompt and displaced lepton reconstruction and identification. The data set and simulated samples used are the same as in the HNL analysis (Chapter 7) and events are selected using the same single lepton triggers. Only MC samples containing ISR and FSR photons associated with a Z boson are considered in this study. Other SM processes have



(a) Feynman diagram of initial state radiation Drell-Yan process where the on-shell photon undergoes external photon conversion with the detector material.

(b) Feynman diagram of final state radiation Drell-Yan process where the on-shell photon undergoes external photon conversion with the detector material.



(c) Feynman diagram of initial state radiation Drell-Yan process where the off-shell photon undergoes internal photon conversion.

(d) Feynman diagram of final state radiation Drell-Yan process where the off-shell photon undergoes internal photon conversion.

Figure 9.9: Feynman diagrams of external (real) and internal (virtual) photon conversions into electron-positron pairs.

negligible cross-sections for photon conversions in comparison to the ones mentioned above and are therefore omitted from this study.

A pair of oppositely-charged, prompt, and isolated muons passing tight identification criteria are selected to reconstruct the Z boson mass, with a similar set of ID, kinematics and quality selections as the leading muons triggered on in the main analysis. The leading muon must exceed $p_T > 26, 29, 26$ GeV per data-taking year while the subleading muon must exceed $p_T > 10$ GeV. Both muons must be within the acceptance of the tracking volume ($|\eta| < 2.4$). The leading muon in the event is tagged as *leading lepton*, ℓ_1 , whereas the subleading one is tagged as *subleading lepton*, ℓ_2 . Electrons from photon conversions are selected in a similar manner as the subleading electrons in the main analysis with $p_T > 5$ GeV and $|\eta| < 2.4$ and passing customized loose identification criteria defined in the main analysis framework. The electron with the highest p_T in the event is selected as a candidate nonprompt electron for the identification efficiency studies and is tagged as *trailing lepton*, e .

A set of selection requirements is imposed in order to define a control region where electrons from asymmetric photon conversions are studied and identification scale factors will be established. Since this control region requires the presence of at least

three leptons, it is called the *3-lepton CR*.

First, a set of generator-matched studies was performed in order to ensure maximal purity in the control region of selected asymmetric photon conversion events. The main concern of the latter is the contamination from fake electrons resulting from multijet production. Maximizing the purity of genuine photon conversions in the selected samples ensures better accuracy in the calculation of scale factors on displaced electrons. Using the generator-matched information on the trailing electron in the two MC sets employed for this study, signal electrons that likely result from prompt photon conversion are flagged accordingly. Electrons that are not matched to a generator-level prompt photon are flagged as background electrons. The following distributions present a justification for the kinematic phase space selection defined in the control region of the asymmetric photon conversion studies. In Fig. 9.10 no phase space preselection cuts are applied, aiming to find the optimal areas enriched in genuine photon conversions.

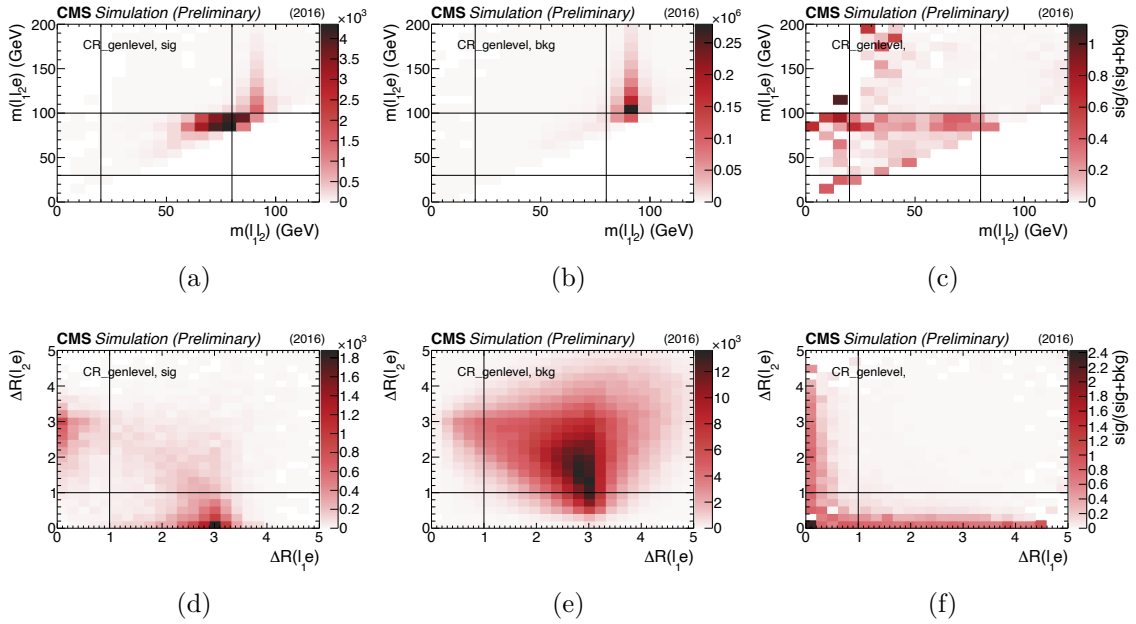


Figure 9.10: Generator-matched distributions aiming to select the optimal phase space enriched in genuine photon conversions in order to maximize purity in the control region used in the asymmetric photon conversion study, using simulation from 2016. The top row shows the trilepton mass vs dimuon mass distributions for signal (left), background (middle) and signal purity (right) highlighting the region of optimal purity selection. The bottom row presents the opening angle $\Delta R(\ell_1 e)$ between the leading muon and trailing electron vs ΔR between the subleading muon and trailing electron.

Figs. 9.10a-9.10c show signal and background distributions between the trilepton mass vs dimuon mass, as well as the signal purity in that kinematic phase space. The optimal selection for enriched photon conversions is found in the $30 < m(\ell_1 \ell_2 e) <$

100 GeV and $20 < m(\ell_1\ell_2) < 80$ GeV regions. This selection also aims to mainly capture FSR processes and reduce potential background from misidentified ISR jets as photons. Figs. 9.10d-9.10f present the signal and background distributions on the ΔR opening angle between the trailing electron and leading/subleading muons, respectively. The signal purity in this kinematic phase space proves that the trailing electron from asymmetric photon conversion is produced closely in angle either to the leading or subleading muon, which may be easily explained by the kinematics of FSR photons in Fig 9.9b. Therefore, the optimal selection was found to be in the $\Delta R(\ell_1 e) < 1$ or $\Delta R(\ell_2 e) < 1$ region. In some bins, the signal purity takes on unphysical values due to the presence of events with negative generator weights since this study uses NLO MC samples. This effect is especially pronounced where event counts are low.

Consequently, the following set of selection criteria is applied in the study of asymmetric photon conversions in the 3-lepton CR, where the first set of cuts has been established based on generator-matched studies using information on the parent of the selected trailing electron:

- Cuts specific to the asymmetric photon conversion study:
 - $30 < m(\ell_1\ell_2 e) < 100$ GeV;
 - $20 < m(\ell_1\ell_2) < 80$ GeV;
 - $\Delta R(\ell_1 e) < 1$ or $\Delta R(\ell_2 e) < 1$;
 - $p_T(\ell_2) > 10$ GeV
- Cuts related to the event pre-selection criteria used in the HNL analysis framework:
 - $p_T(\ell_1) > 26, 29, 26$ GeV (2016, 2017, 2018);
 - $p_T(e) > 5$ GeV;
 - $|\eta(\ell_{1,2})| < 2.4$, $|\eta(e)| < 2.4$;
 - $n_{\text{jets}} > 0$;
 - $p_T(\text{boosted jet}) > 30$ GeV, $|\eta(\text{boosted jet})| < 2.4$ and relative isolation of the electron $I_{\text{rel}}(e) > 0.15$;
 - $p_T(\text{resolved jet}) > 20$ GeV, $|\eta(\text{boosted jet})| < 2.4$.

where *boosted jet* is defined as $\Delta R(\text{jet}, e) < 0.4$ and *resolved jet* is separated from the trailing electron, $\Delta R(\text{jet}, e) > 0.4$.

Two sets of data-to-MC plots are made based on whether the trailing electrons are simply reconstructed as PF electrons or if they additionally pass custom loose electron identification criteria. While the former serves as the denominator in the subsequent efficiency calculations, the latter constitutes the numerator as per Equation 9.1. Figures 9.11 and 9.13 present a broad set of data to simulation plots for a number of kinematic observables for the 2016 data-taking year. Further such plots are shown in Appendix E.1, Figs. E.1-E.5. Overall, good agreement between data and MC is observed in most such variables. Electron ID efficiencies are evaluated for a subset of them ($d_{xy}^{sig}(e)$, $d_{xy}(e)$, $p_T(e)$, $|\eta(e)|$, dilepton and trilepton masses) by dividing the right-hand side plots (ID+reco) with the left-hand side ones (reco only), as presented in Fig. 9.14. Statistical uncertainties are evaluated for both efficiency distributions. Uncertainties on data efficiency are calculated via the Clopper-Pearson binomial proportion confidence interval at 1σ deviation, using Eq. 5.15 [138]. Since MC-simulated samples at NLO contain negative weights, the Clopper-Pearson confidence interval is not suitable for calculating uncertainties on MC efficiencies. Consequently, statistical uncertainties on the MC efficiencies are evaluated via error propagation. For more details on the statistical estimates used in the evaluation of uncertainties, please consult Sec. 5.3.3.

As a comparison to the evaluated 1D MC efficiency distributions on electrons from asymmetric photon conversions, a set of MC efficiency distributions was produced on reconstructed signal electrons from 3 benchmark HNL decays, using custom electron identification. Fig. 9.16 shows that reconstructed signal electrons have higher MC efficiencies than displaced electrons captured via asymmetric photon conversions. Differences between electron efficiencies from the 3 HNL benchmark models may occur whether the reconstructed signal electron is boosted inside the HNL jet or not. Since lighter HNL models tend to produce highly boosted decay particles, the daughter electron would be merged inside the HNL jet, having an environment with higher activity and lower isolation. The blue distributions in Fig. 9.16 correspond to a boosted HNL scenario, whereas the red and green distributions may be resolved from the HNL jet, hence showing better agreement with each other, as opposed to the blue one with different event topology.

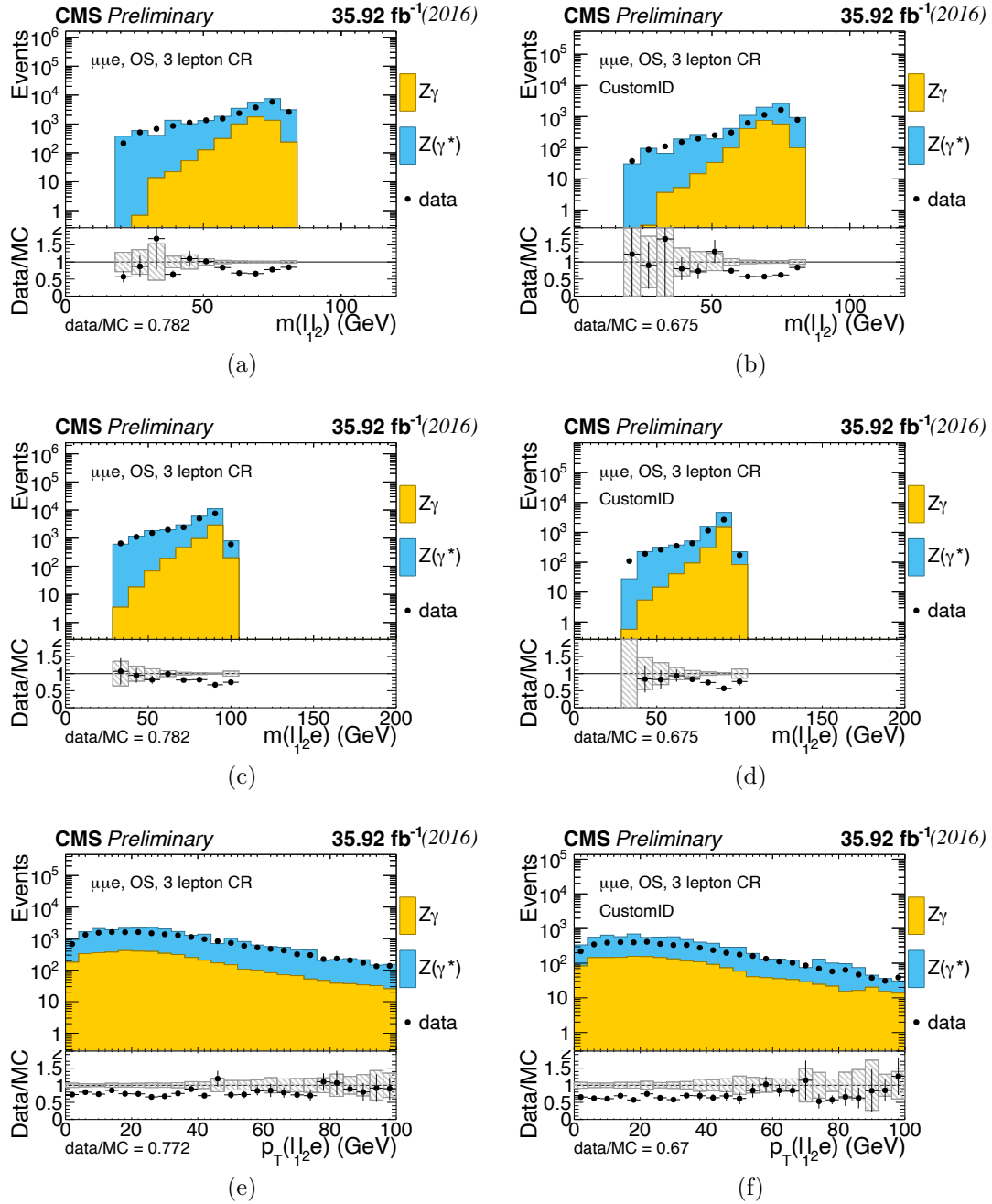


Figure 9.11: Data to MC agreement plots in the 3-lepton CR for dilepton mass $m(l_1 l_2)$, trilepton mass $m(l_1 l_2 e)$ and trilepton p_T . Plots on the left do not have custom electron ID applied whereas those on the right do.

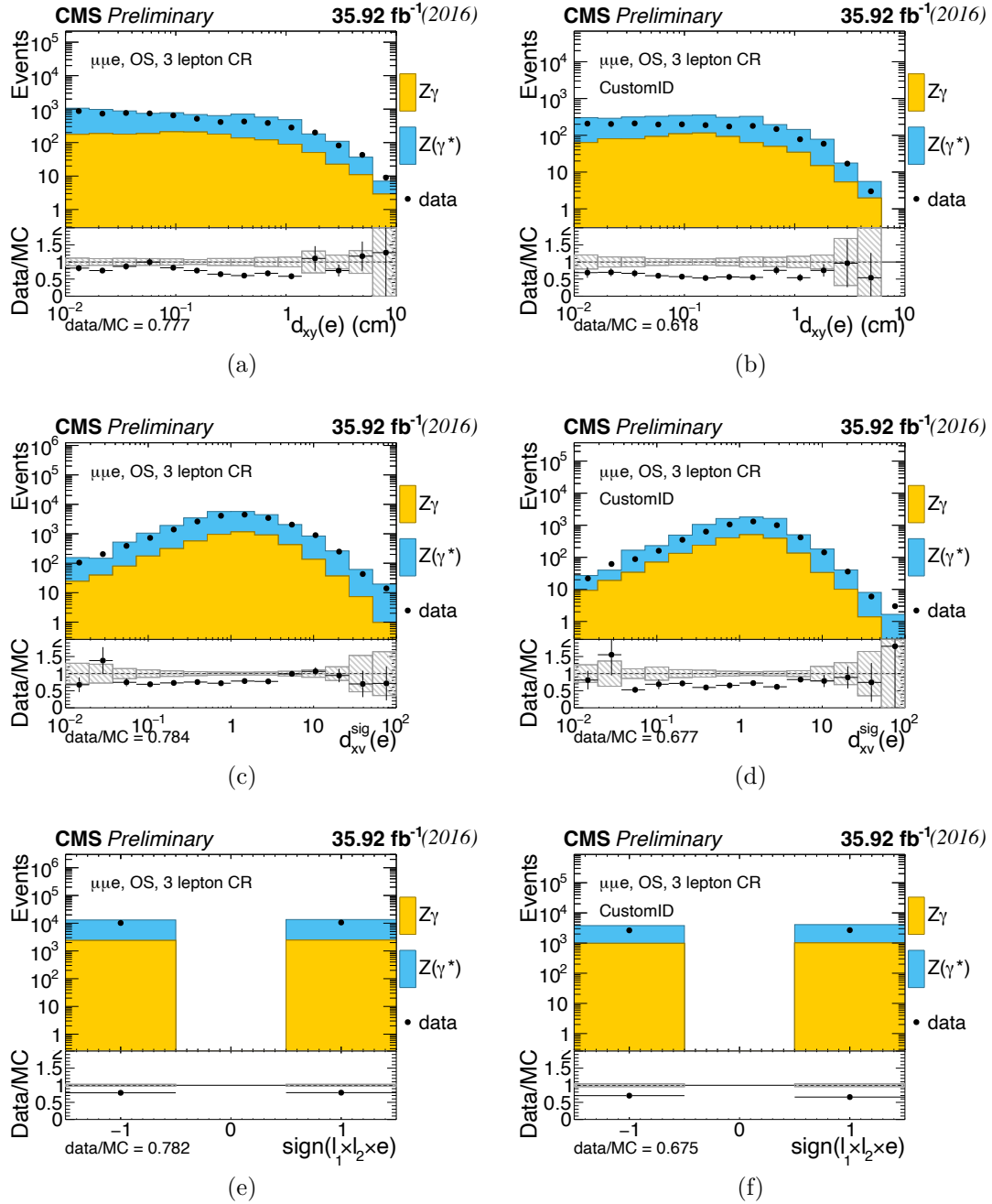


Figure 9.12: Data to MC agreement plots in the 3-lepton CR for the trailing lepton $d_{xy}(e)$, $d_{xy}^{sig}(e)$ and trilepton charge. Plots on the left do not have custom electron ID applied whereas those on the right do.

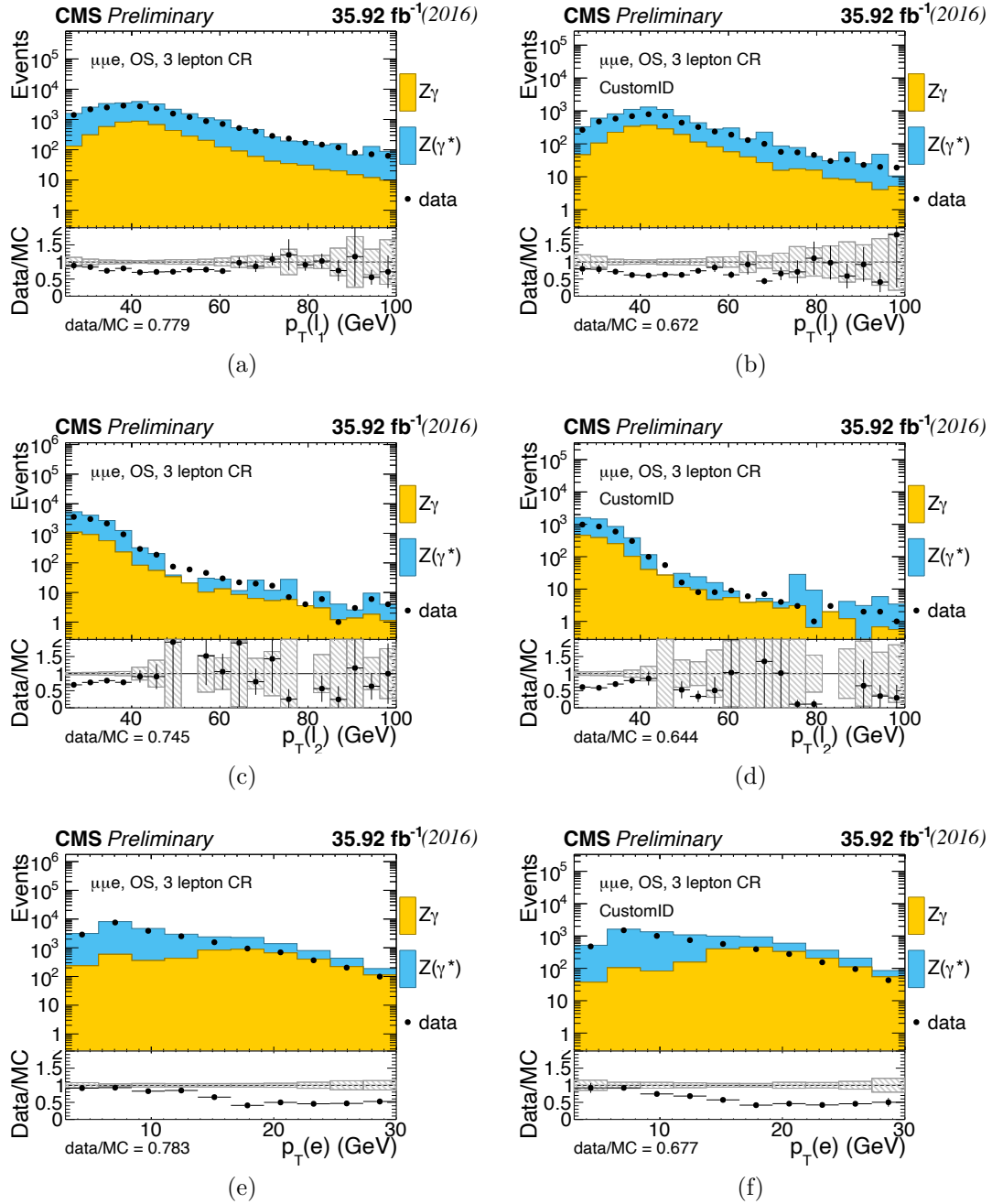


Figure 9.13: Data to MC agreement plots in the 3-lepton CR for the leading, subleading and trailing lepton p_T distributions. Plots on the left do not have custom electron ID applied whereas those on the right do.

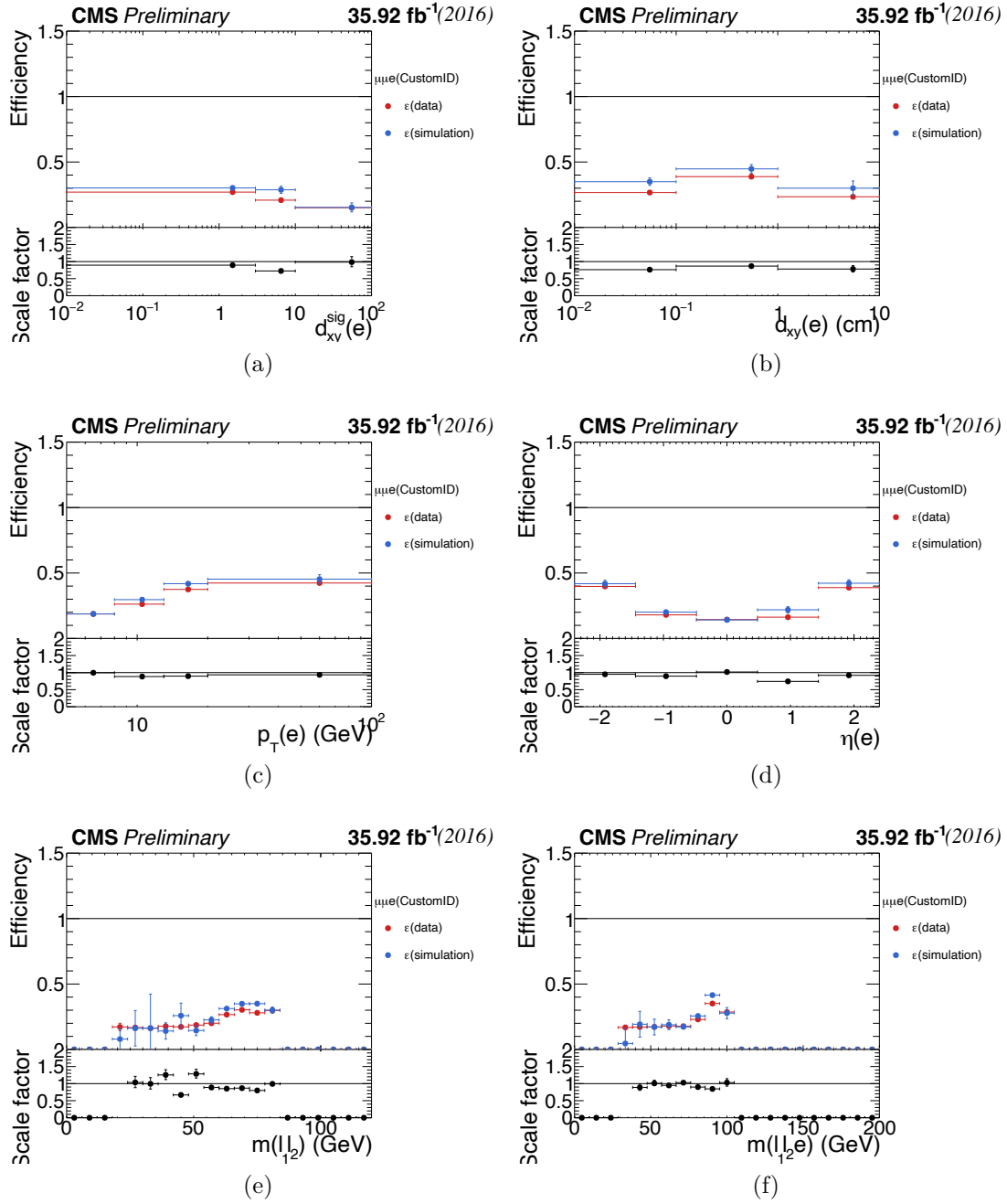


Figure 9.14: Efficiency and scale factor distributions evaluated for displaced custom electron identification performance for a number of kinematic variables of interest (2016).

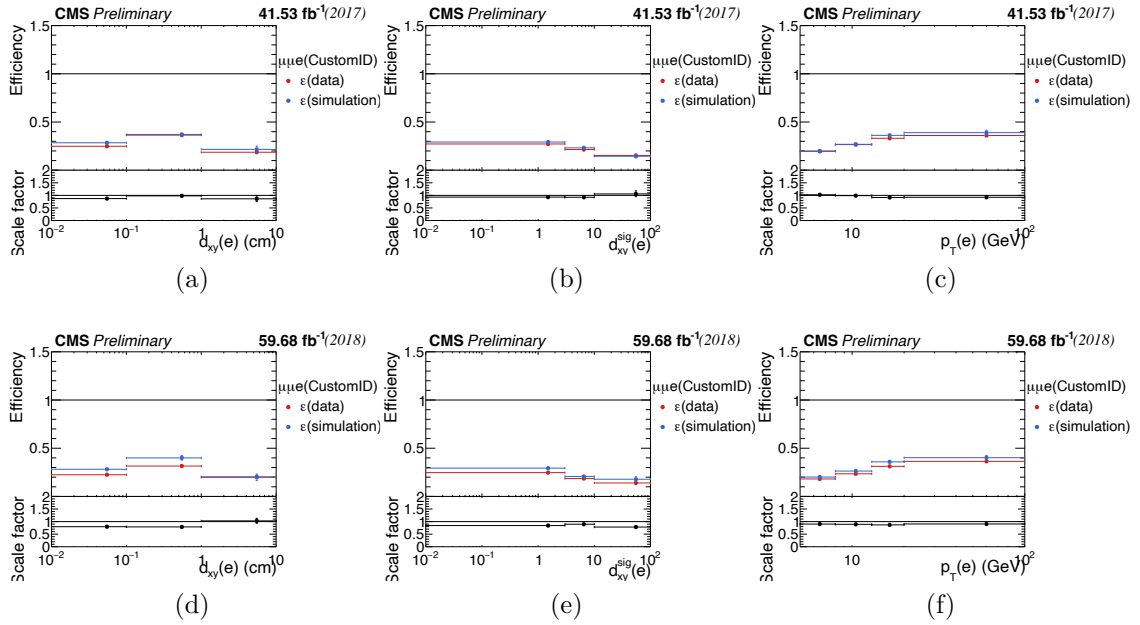


Figure 9.15: 1D efficiency and scale factor distributions evaluated for displaced electron custom identification performance for $d_{xy}(e)$, $d_{xy}^{sig}(e)$ and $p_T(e)$ kinematic variables for 2017 (top) and 2018 (bottom) data-taking years.

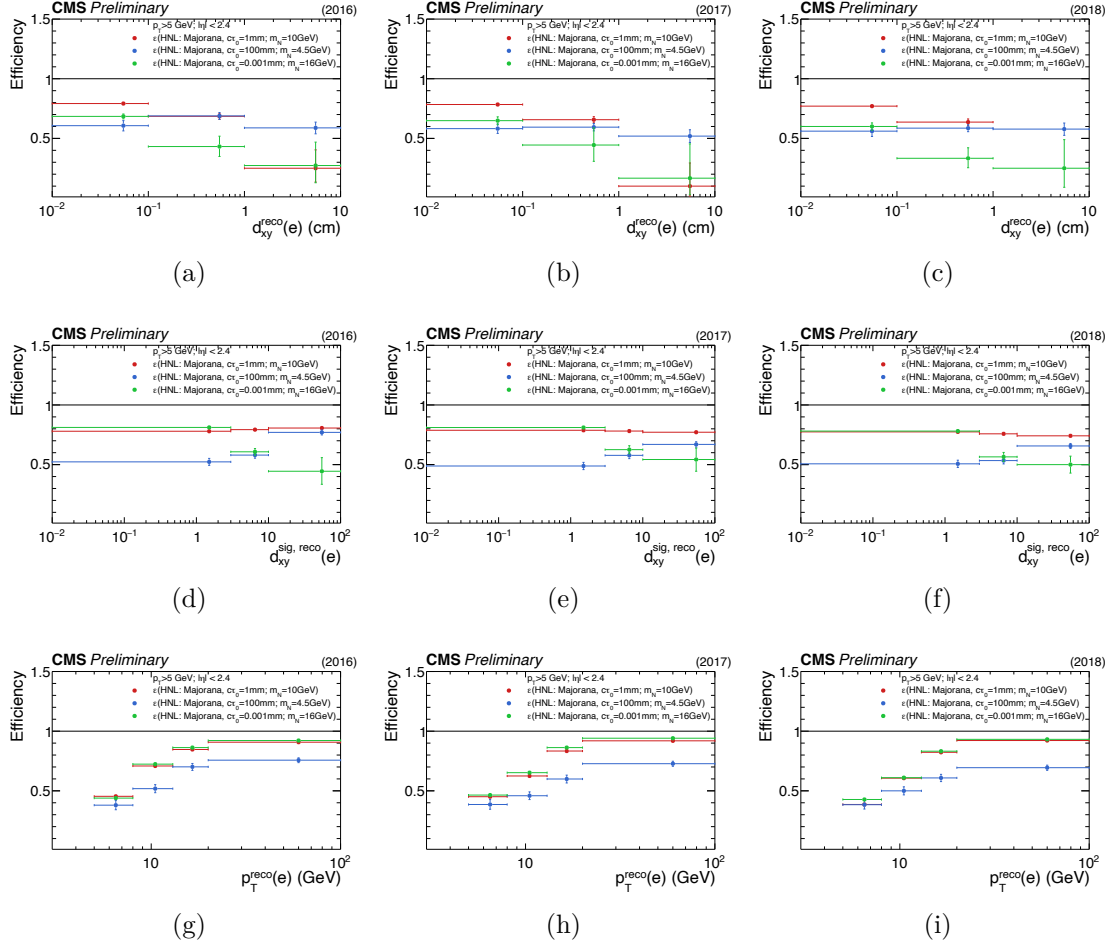


Figure 9.16: 1D efficiency distributions evaluated for reconstructed signal electrons from 3 HNL benchmark models using custom identification. The distributions are shown for $d_{xy}(e)$ (top), $d_{xy}^{sig}(e)$ (middle) and $p_T(e)$ (bottom) kinematic variables in the three data-taking years 2016, 2017 and 2018 (left to right).

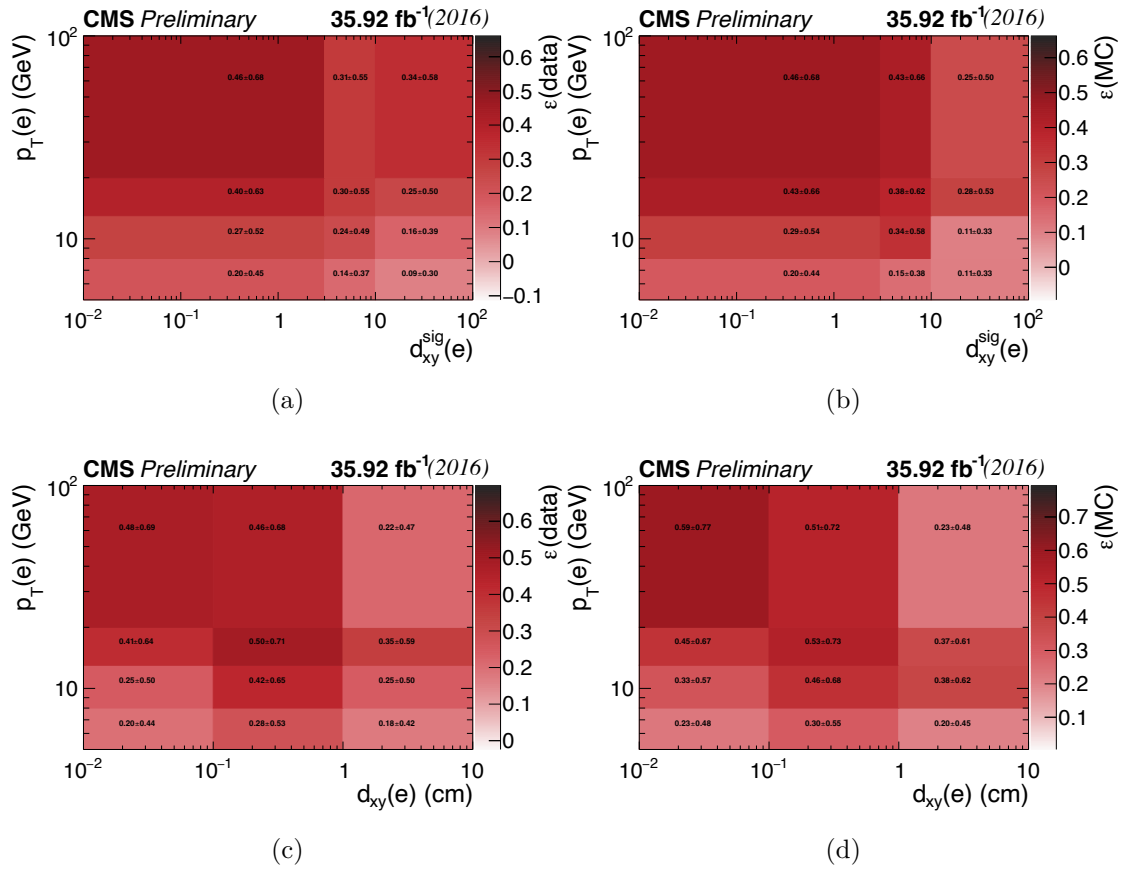


Figure 9.17: 2D efficiency distributions evaluated for displaced electron identification on both data (left) and MC (right) for kinematic variable pairs of interest (2016).

Finally, a set of “raw” scale factors are evaluated for displaced electron identification by dividing the data efficiencies with the MC ones, as those presented in Figs. 9.17. The results on the 2D efficiency distributions presented in Fig. 9.17 are on the 2016 data set, whereas those for 2017 and 2018 can be found in Appendix E.2. Uncertainties are evaluated via error propagation for both up and down deviation separately and their average is taken. Tables 9.5 and 9.6 present the set of scale factors measured for $p_T(e)$ vs $d_{xy}^{sig}(e)$ and $d_{xy}(e)$, respectively, for all three data-taking years on displaced electrons from asymmetric photon conversions (their heatmap versions are shown in Appendix E.3, as well as for all subsequent tables below). Due to low MC statistics in higher displacement $d_{xy}(e)$ bins, the scale factors at higher impact parameter values have large uncertainties.

Table 9.5: Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data-taking years.

(a) CMS Preliminary, 35.92 pb ⁻¹ (2016)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	1.00±0.09	0.71±0.15	1.35±0.59
13.0-20.0	0.93±0.06	0.78±0.10	0.90±0.34
8.0-13.0	0.93±0.07	0.70±0.14	1.46±0.90
5.0-8.0	1.02±0.10	0.96±0.22	0.81±0.37
(b) CMS Preliminary, 41.53 pb ⁻¹ (2017)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.97±0.08	0.85±0.13	1.69±0.90
13.0-20.0	0.95±0.05	0.93±0.10	0.72±0.16
8.0-13.0	0.98±0.06	1.01±0.13	1.52±0.55
5.0-8.0	1.04±0.07	1.02±0.16	1.03±0.38
(c) CMS Preliminary, 59.68 pb ⁻¹ (2018)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.93±0.07	0.97±0.14	0.94±0.32
13.0-20.0	0.89±0.04	0.93±0.09	0.72±0.14
8.0-13.0	0.92±0.06	0.89±0.12	0.61±0.21
5.0-8.0	0.88±0.06	1.12±0.15	1.47±0.49

Table 9.6: Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}(e)$ kinematic variable pair for all three data-taking years.

(a) CMS Preliminary, 35.92 pb ⁻¹ (2016)			
p_T (GeV) vs $d_{xy}(e)$ (cm)	0.01-0.1	0.1-1.0	1.0-10.0
20.0-100.0	0.81±0.12	0.90±0.14	0.95±0.65
13.0-20.0	0.90±0.11	0.94±0.11	0.95±0.26
8.0-13.0	0.76±0.15	0.92±0.16	0.64±0.20
5.0-8.0	0.86±0.18	0.93±0.17	0.86±0.40

(b) CMS Preliminary, 41.53 pb ⁻¹ (2017)			
p_T (GeV) vs $d_{xy}(e)$ (cm)	0.01-0.1	0.1-1.0	1.0-10.0
20.0-100.0	0.95±0.17	1.15±0.18	0.50±0.26
13.0-20.0	0.88±0.09	0.94±0.09	0.83±0.31
8.0-13.0	0.92±0.13	1.05±0.15	0.97±0.29
5.0-8.0	1.07±0.20	1.05±0.16	1.03±0.38

(c) CMS Preliminary, 59.68 pb ⁻¹ (2018)			
p_T (GeV) vs $d_{xy}(e)$ (cm)	0.01-0.1	0.1-1.0	1.0-10.0
20.0-100.0	1.00±0.14	0.90±0.12	1.57±0.77
13.0-20.0	0.90±0.09	0.85±0.08	1.29±0.41
8.0-13.0	0.84±0.14	0.84±0.10	1.02±0.25
5.0-8.0	0.89±0.14	0.84±0.14	0.86±0.35

Since the scale factor results presented above may lack robustness due to statistical limitations in some histogram bins, we introduce a “geometric mean” approach to determining synthetic 2D scale factor histograms extrapolated from single-variable scale factors (See bottom of Figs. 9.14a, 9.14c and 9.15). An outer product is performed between the 1D scale factor histograms of the $d_{xy}^{sig}(e)$ and $p_T(e)$ kinematic variables and a square root is applied on the resulting 2D distribution, i.e.:

$$SF(i, j) = \sqrt{SF(i)SF(j)} \quad (9.5)$$

$$\sigma(SF(i, j)) = \frac{1}{2} \sqrt{\left(SF(i) \frac{\sigma^2(SF(j))}{SF(j)} + SF(j) \frac{\sigma^2(SF(i))}{SF(i)} \right)} \quad (9.6)$$

where $\sigma(SF(i, j))$ is the uncertainty on the 2D scale factor calculated via error propagation in bin (i, j) of the 2D distribution. The resulting distributions are presented in Table 9.7. Most bins have lower uncertainties and at higher displacement show more robust values in comparison to the “raw” scale factors presented in Table 9.5. In addition, pull histograms are evaluated between the two approaches

as presented in Table 9.8, showing that most scale factor values are within 1–1.5 standard deviations between the two methods, providing confidence in the “geometric mean” approach. Mean pull values per 2D distribution estimated for each data-taking year are 2016: 0.16 ± 0.57 ; 2017: 0.21 ± 0.65 ; 2018: 0.39 ± 0.72 showing a relatively low bias for the “geometric mean” estimator. In addition, a set of scale factors is evaluated similarly for the $d_{xy}(e)$ vs $p_T(e)$, as well as the $\eta(e)$ vs $p_T(e)$ variable pairs shown in Appendix E.3 for all data-taking years.

Table 9.7: Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data-taking years. Uncertainties presented are statistical only.

(a) CMS Preliminary, 35.92 pb ⁻¹ (2016)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.91±0.03	0.82±0.03	0.96±0.08
13.0-20.0	0.90±0.02	0.81±0.02	0.94±0.07
8.0-13.0	0.89±0.02	0.80±0.02	0.93±0.07
5.0-8.0	0.94±0.02	0.85±0.02	0.99±0.08
(b) CMS Preliminary, 41.53 pb ⁻¹ (2017)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.93±0.03	0.92±0.03	0.99±0.06
13.0-20.0	0.92±0.02	0.92±0.03	0.98±0.06
8.0-13.0	0.96±0.02	0.96±0.03	1.02±0.06
5.0-8.0	0.98±0.02	0.97±0.03	1.04±0.06
(c) CMS Preliminary, 59.68 pb ⁻¹ (2018)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.87±0.02	0.90±0.03	0.84±0.04
13.0-20.0	0.86±0.01	0.88±0.02	0.83±0.04
8.0-13.0	0.87±0.01	0.89±0.02	0.84±0.04
5.0-8.0	0.87±0.01	0.90±0.02	0.84±0.04

Table 9.8: Pull values estimated between the “raw” and “geometric mean” scale factor approaches on the $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data-taking years.

(a) CMS Preliminary, 35.92 pb ⁻¹ (2016)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.88	-0.76	0.65
13.0-20.0	0.51	-0.28	-0.12
8.0-13.0	0.49	-0.73	0.59
5.0-8.0	0.73	0.50	-0.46

(b) CMS Preliminary, 41.53 pb ⁻¹ (2017)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.45	-0.58	0.78
13.0-20.0	0.55	0.10	-1.52
8.0-13.0	0.30	0.39	0.90
5.0-8.0	0.85	0.27	-0.03

(c) CMS Preliminary, 59.68 pb ⁻¹ (2018)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.77	0.48	0.29
13.0-20.0	0.67	0.48	-0.74
8.0-13.0	0.94	-0.04	-1.10
5.0-8.0	0.22	1.49	1.27

Since the results from the “geometric mean” approach show robust values with lower uncertainties, while still being compatible with the scale factors calculated via the “raw” approach, the former are propagated to the final limits. In addition to the statistical uncertainties evaluated above via error propagation, a set of systematic uncertainties was added to the final scale factor uncertainties.

Asymmetric photon conversion events are divided by the number of pile-up primary vertices in each event into *high pile-up* and *low pile-up* categories, based on Fig. 3.3. Once scale factors are evaluated for both categories, half the difference between the low and high pile-up region scale factors is applied as systematic uncertainty to the scale factors derived inclusively (Table 9.7). The total uncertainty on the scale factors propagated in the analysis is the quadrature sum of their statistical and systematic uncertainties:

$$\sigma_{\text{total}} = \sqrt{\sigma_{\text{statistic}}^2 + \sigma_{\text{systematic}}^2}, \quad (9.7)$$

presented in Table 9.9. Scale factors and efficiencies evaluated in the low and high pile-up regions are shown in Appendix E.4. Pull histograms are produced between

the inclusive scale factor values and those evaluated in the low and high pile-up regions, showing that most scale factor values are within 1–1.5 standard deviations. Mean pull values per 2D distribution on average show bias towards higher scale factor values in both the low and high pile-up regions with standard deviation per distribution ranging between 0.11-0.91.

Table 9.9: Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data-taking years including systematic uncertainties. These scale factors are eventually propagated further in the HNL analysis.

(a) CMS Preliminary, 35.92 pb ⁻¹ (2016)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.91±0.05	0.82±0.08	0.96±0.17
13.0-20.0	0.90±0.05	0.81±0.08	0.94±0.16
8.0-13.0	0.89±0.07	0.80±0.10	0.93±0.18
5.0-8.0	0.94±0.06	0.85±0.09	0.99±0.18
(b) CMS Preliminary, 41.53 pb ⁻¹ (2017)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.93±0.06	0.92±0.06	0.99±0.12
13.0-20.0	0.92±0.05	0.92±0.05	0.98±0.13
8.0-13.0	0.96±0.04	0.96±0.04	1.02±0.15
5.0-8.0	0.98±0.08	0.97±0.08	1.04±0.11
(c) CMS Preliminary, 59.68 pb ⁻¹ (2018)			
p_T (GeV) vs $d_{xy}^{sig}(e)$	0.01-3.0	3.0-10.0	10.0-100.0
20.0-100.0	0.87±0.02	0.90±0.04	0.84±0.04
13.0-20.0	0.86±0.03	0.88±0.02	0.83±0.04
8.0-13.0	0.87±0.02	0.89±0.03	0.84±0.04
5.0-8.0	0.87±0.02	0.90±0.05	0.84±0.05

9.4 Summary

This Chapter presented the main sources of systematic uncertainty considered in the search for long-lived HNL particles. A special focus was given to the strategies used in evaluating the efficiencies and associated uncertainties of displaced leptons, eventually assigned as corrections to signal processes. Scale factors assigned for muon identification criteria (Fig. D.1) prove a robust performance of the loose muon ID used in the HNL search. Muon reconstruction scale factors (Fig. D.3) also show a solid performance of the CMS detector in the majority of the kinematic phase space considered, except at low- p_T . The outstanding performance on muon reconstruction and identification of the CMS detector is attributed to the additional information used from the muon chambers, which the electron reconstruction and identification lacks. Therefore, as expected, electron identification efficiencies (Fig. E.15) showed a considerably lower performance to muons. Nevertheless, the estimated scale factors are within 10-20% from unity.

The majority of the systematic uncertainties discussed in this Chapter affect the predicted HNL signal yield since it was taken from simulation, whereas the yield of background contribution is taken from data directly. The subject of Chapter 10 is a detailed discussion on the methods used in estimating the expected background contributions in the HNL search, as well as the treatment of its systematic uncertainties.

Chapter 10

Background estimation and results

This Chapter presents the data-driven background estimation approach employed in predicting the yield of SM background events in the blinded SR of the long-lived HNL search. Using this information along with the expected HNL signal yields, the background-only hypothesis was tested via goodness-of-fit tests. No excess in data was found, therefore upper limits were set on the signal strength in the HNL mass, m_N , and coupling, V_ℓ^2 , parameter space.

10.1 Background estimation

This Section briefly presents the data-driven SM background estimation in the blinded SR of the analysis via an ABCD technique. The definition of the discriminating variables employed, as well as tests of the validity of the ABCD method closure in two validation regions, and goodness-of-fit likelihood tests are given.

The methods and results presented in this Section are based on the work of an analysis team the author of this Thesis was part of and, made contributions to. Nevertheless, certain figures and ideas in this Section are based on work previously published in Ref. [109].

10.1.1 Introduction

The long-lived HNL search presented in this Thesis employs a data-driven background estimation technique relying on the ABCD method to evaluate the expected background event yields in the blinded SR. The principles of the ABCD technique are

discussed in Sec. 5.3.6, introducing the need for two statistically independent variables that possess discriminating power between signal and background events. The two variables used in the ABCD background estimation method of the HNL analysis are the displaced jet tagger information and the event-level BDT score. As already emphasised in Chapter 8, the event-level BDT was designed to be uncorrelated from the DNN jet tagger, proved via low MI scores in the 2D heatmap distributions between the two variables (Fig. 8.11) for a number of dilepton flavour categories on both signal and background distributions. Even though the independence requirement between two discriminating variables may not hold for mixed background sources, via extensive tests (Ref. [109]), as well as their combined distribution in Fig. 8.11 provides confidence in the validity of the ABCD method. A preselection constraint is placed on the $m(\ell_1, \ell_2, j^*)$ variable, selecting events with $70 \text{ GeV} < m(\ell_1, \ell_2, j^*) < 90 \text{ GeV}$ before performing the ABCD background estimation method. This preselection constraint ensures increased purity of signal events in the blinded SR.

10.1.2 Threshold optimisation

The lower threshold of the tagger score, P_{low} , was increased from its default value of 0 in the ABCD method and optimised for merged and resolved scenarios (Ref. [109]). Increasing the lower tagger threshold increases the fraction of selected signal-like events and moves away from the bulk of the background distribution. The drawback of raising P_{low} is the reduction of statistical precision in the ABCD method. Separate P_{low} thresholds are set for the merged and resolved event categories since the tagger performance significantly differs for the two event scenarios. The optimised lower tagger threshold values are $P_{low} = 0.3$ for merged and $P_{low} = 0.2$ for resolved event categories.

In addition, the mid-boundaries of the discriminating variables in the ABCD regions were optimised, as well, taking into account several benchmark HNL scenarios in order to avoid bias towards certain HNL model parameters. Different mid-boundary thresholds were applied for individual event categories and data-taking years, as the performance of the two discriminating variables, as well as the yields of background distributions and data-taking conditions vary. The thresholds were optimised based on the discovery significance (Eq. 5.36), requiring a minimum of 10 background event counts in the A, B and C sidebands. The BDT_{mid} values range between 0.5-0.8, whereas P_{mid} range between 0.6-0.8 for merged and 0.4-0.5 for resolved categories, respectively. The phase-space division of the ABCD background estimation method is schematically illustrated in Fig. 10.1.

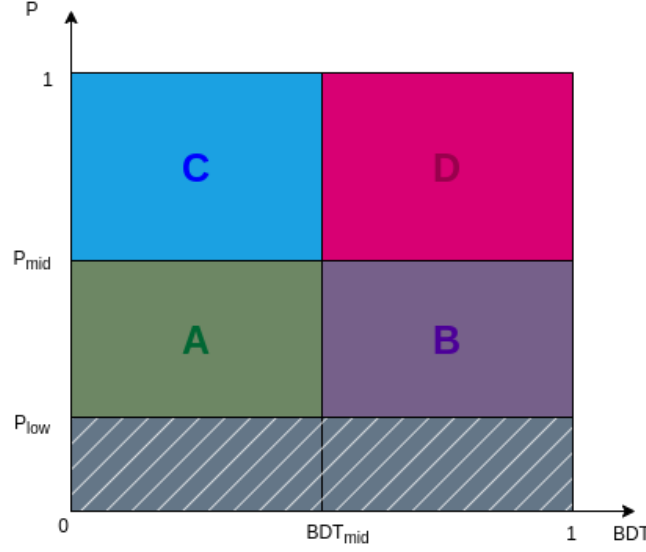


Figure 10.1: Schematic diagram of the ABCD background estimation technique definition applied for the long-lived HNL search, where D is the blinded SR of the analysis. The low tagger score, P_{low} , region (shaded) is removed to maximise signal-like jets being selected.

10.1.3 Closure tests

To test the assumptions of the ABCD method, so-called closure tests were performed in signal-depleted validation regions (VR). These tests aim to compare the predicted yield, $\tilde{N}_D$, with the true yield, N_D , in the blinded region of the VR. The section below briefly presents the results obtained in two closure tests performed on data, as detailed in Ref. [109]. The two VRs of the closure tests were defined as:

- **VR1 (mass sideband):** is defined by inverting the $m(\ell_1, \ell_2, j^*)$ preselection criteria to capture the low-mass sideband region, $m(\ell_1, \ell_2, j^*) < 70$ GeV. The mid-tagger threshold, P_{mid} is reduced in order to increase the statistical significance in the "blinded region". Fig. 10.2 shows the outcome of the closure test in VR1, providing confidence in the validity of the ABCD method in this VR.
- **VR2 (tagger sideband):** is defined by subdividing the low tagger score sideband $P \in [P_{\text{low}}, P_{\text{mid}}]$ into four new regions such that $AB \rightarrow A'B'C'D'$. Fig. 10.3 shows the outcome of the closure test in VR2.

The κ parameters displayed in the two sets of closure tests is a parameter introduced to quantify the extent of bias in the ABCD method:

$$\tilde{N}_D^{\text{bkg}} = \kappa N_B^{\text{bkg}} \times N_C^{\text{bkg}} / N_A^{\text{bkg}}. \quad (10.1)$$

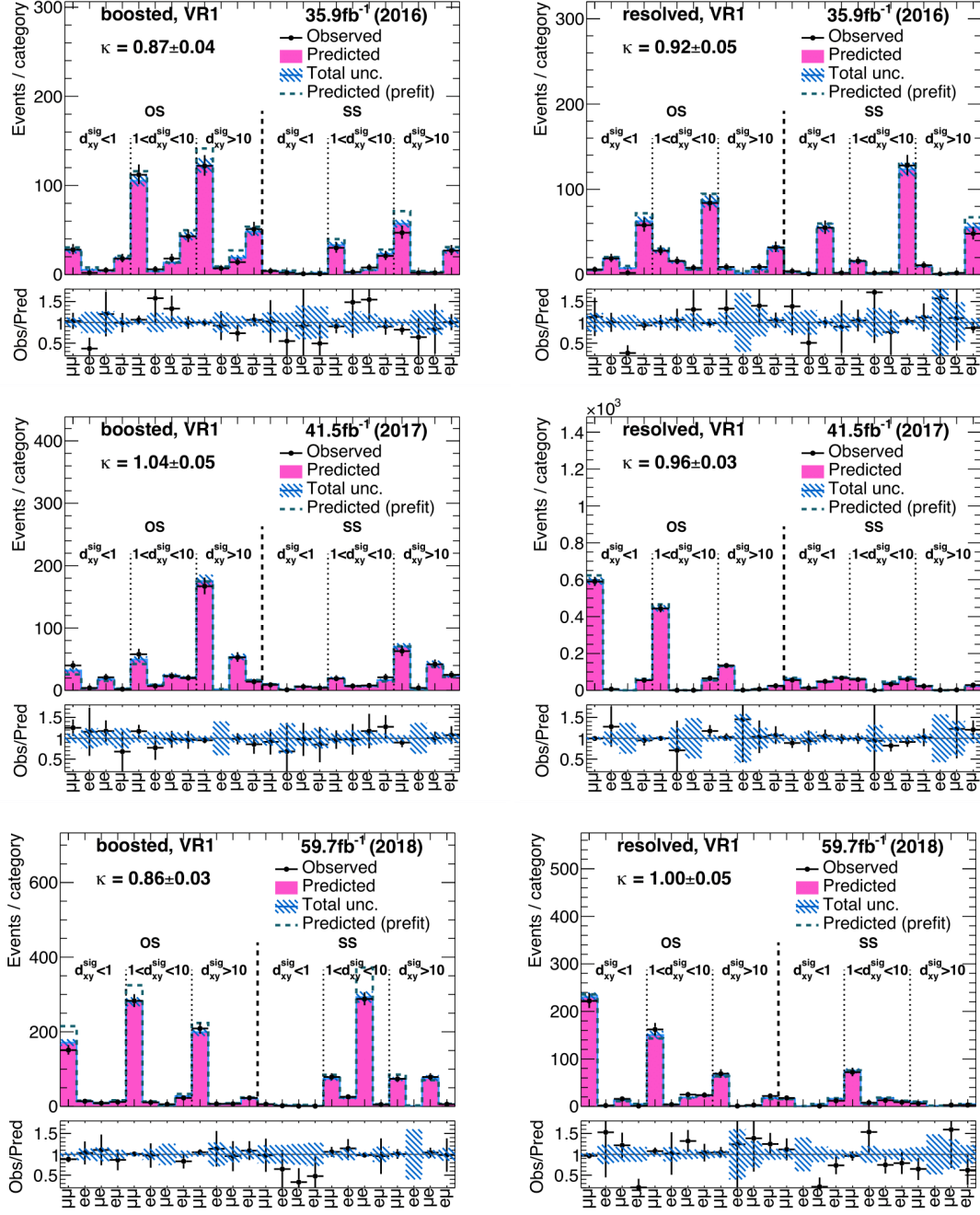


Figure 10.2: Outcome of the closure test performed in the low-mass region VR1 for merged (left) and resolved (right) event categories for all three data-taking years (2016-2018 from top to bottom).

In total, six κ parameters were constrained by data-taking year and event topology (merged vs resolved), reflecting the different data-taking conditions and tagger performance on jet topologies. A maximum likelihood fit was performed for each set of 24 event categories within a macrocategory, the best-fit value being shown in the closure test result plots. The κ parameters corresponding to the closure tests are consistent with unity within uncertainties, providing further confidence in the ABCD method. A conservative 15% systematic uncertainty was assigned for each of

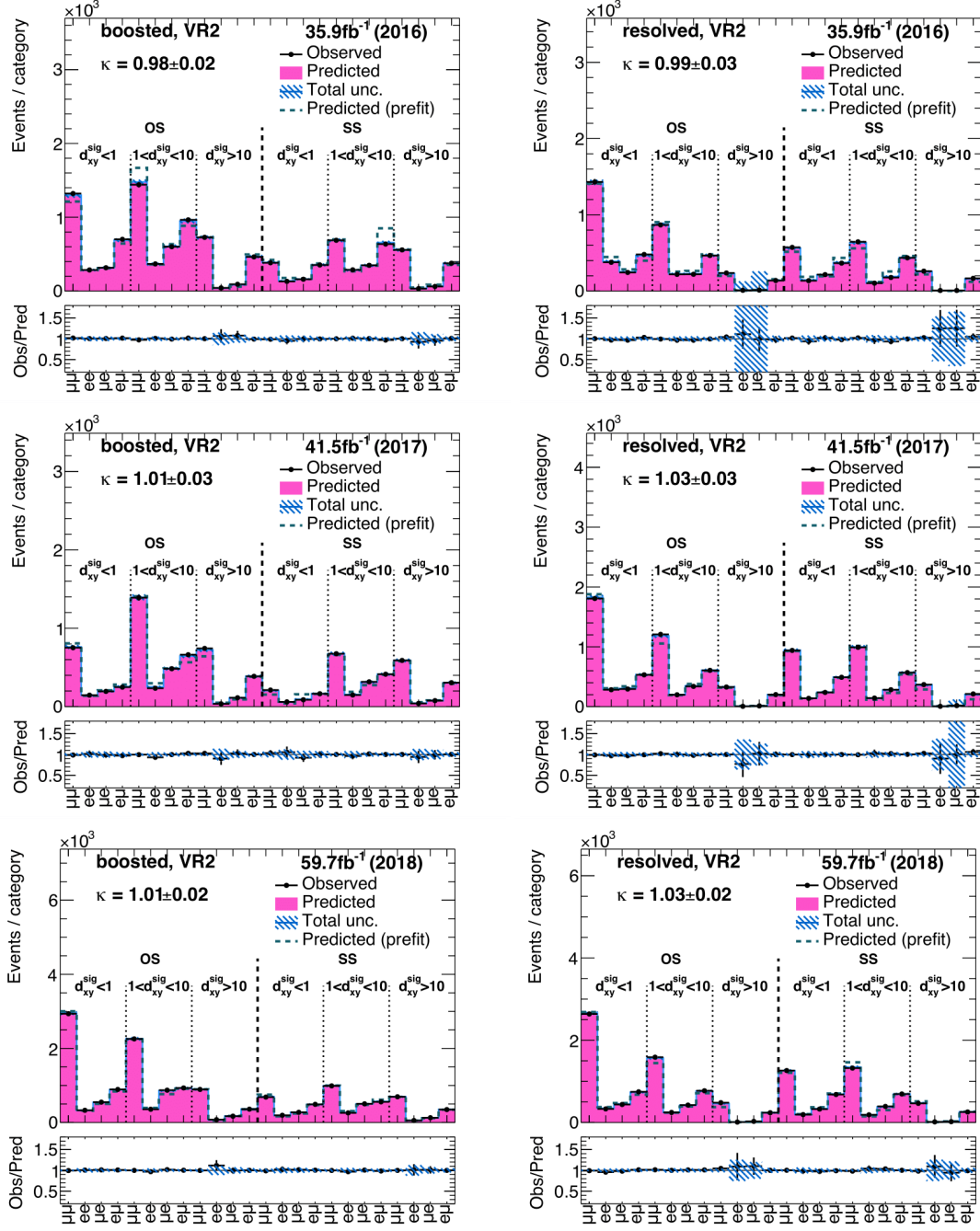


Figure 10.3: Outcome of the closure test performed in the tagger sideband VR2 for merged (left) and resolved (right) event categories for all three data-taking years (2016-2018 from top to bottom).

the six macrocategories, covering potential non-closure issues beyond the precision of the closure tests.

Finally, the ABCD background estimation method was further validated by performing GOF tests in both VRs, according to Eq. 5.35. The saturated likelihood test statistic distributions were generated with MC toys, asymptotically reaching a χ^2 distribution at large generated samples. The predicted probability distributions were

compared to the measured GOF value in data, deriving a corresponding p -value to the results, as presented in Fig. 10.4. The high observed p -values in the two closure tests, provide further confidence in the robustness of the ABCD method.

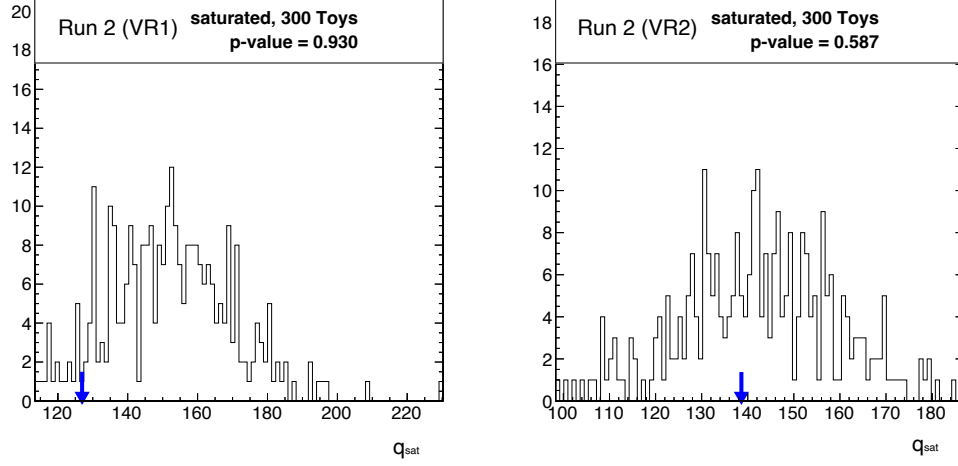


Figure 10.4: The predicted saturated likelihood test statistic distributions generated with toy-data and the observed value in data (blue arrow) for both the VR1 (left) and VR2 (right) GOF tests.

10.2 Results

In the extensive validation studies on the reliability of the ABCD background estimation method presented in the previous Section, the blinded region D of the SR (see Fig. 10.1) was unblinded, revealing the observed data yields in each category, as shown in Fig. 10.5. The yields for the expected limits were obtained from a fit to the Asimov dataset. No excess in data was found, the observed and expected background yields are in agreement. Expected yields for two representative benchmark HNL models are overlaid, and expected to be excluded by the analysis. A GOF test was performed proving consistency with the background-only hypothesis in the unblinded data, observing a large p -value for the test statistic (see Fig. 10.6).

10.2.1 Limits in mass-coupling plane

As no evidence for long-lived HNLs was found in the SR of the analysis, upper-limits (ULs) on the HNL production cross-sections at 95% confidence level (CL), $\sigma_{\text{UL}}(m_N, V_\ell^2)$, were determined in the mass-coupling space for various HNL benchmark scenarios. The observed and expected HNL signal cross-sections were compared to the theoretical ones ($\sigma_{\text{th}}(m_N, V_\ell^2)$), evaluated by MADGRAPH at LO precision in

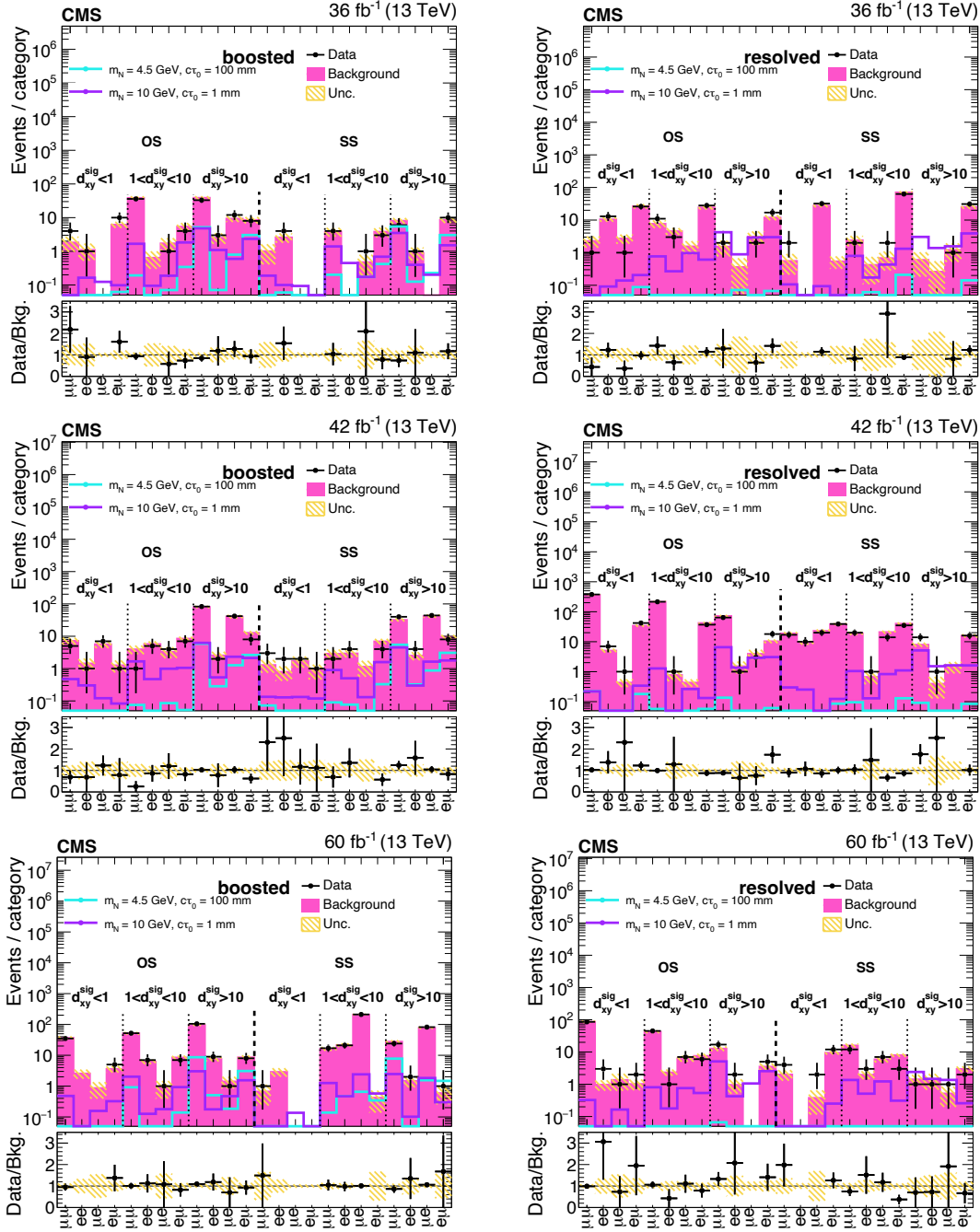


Figure 10.5: Observed and predicted yields in the SR for merged (left) and resolved (right) event categories for all three data-taking years (2016-2018 from top to bottom). Expected yields for two benchmark HNL models are overlaid.

associated pre-showering and scaled by a k-factor of 1.1 to be consistent with NLO predictions (as discussed in Sec. 7.1.1). A given HNL benchmark model is excluded if its signal strength is lower than the predicted theoretical one, such that:

$$\mu(m_N, V_\ell^2) = \frac{\sigma_{UL}(m_N, V_\ell^2)}{\sigma_{th}(m_N, V_\ell^2)} < 1, \quad (10.2)$$

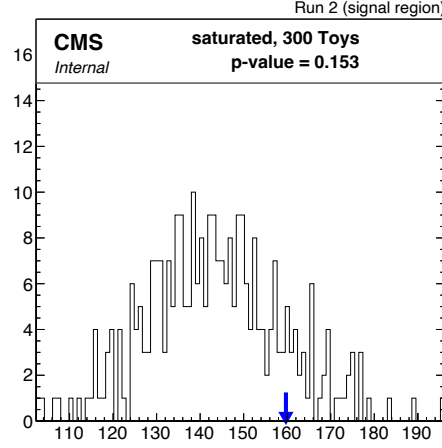


Figure 10.6: The predicted saturated likelihood test statistic distribution generated with toy-data and the observed value (blue arrow) for the SR GOF test.

where $\mu(m_N, V_\ell^2)$ is the signal strength parameter. Performing this test on all available HNL signal scenarios, a cubic spline interpolated contour is drawn at the boundaries of HNL benchmark models that are excluded, as well as marking the 1 and 2 sigma uncertainty on the exclusion limit. Figure 10.7 shows the obtained exclusion limits for both Dirac and Majorana HNL models for a number of coupling scenarios, combining results from all three data-taking years.

10.2.2 Discussion of the results

A broad parameter space in mass and coupling values was excluded by the long-lived HNL search presented in this Thesis, including mixed-coupling scenarios involving all three lepton generations. Sensitivity to prompt HNL scenarios was maintained by the displaced jet tagger not explicitly requiring secondary vertices, therefore events with small displacement could be retained. Best sensitivity is obtained for pure-muon coupling scenarios due to the excellent reconstruction and identification capabilities of the CMS detector. Electron reconstruction, particularly for displaced scenarios is less efficient in the CMS experiment, therefore pure-electron coupling events have slightly reduced limits. As discussed in Sec. 7.1.1, triggering on tau leptons was rather inefficient, therefore the analysis is not sensitive to pure-tau couplings. Nevertheless, it was still possible to measure mixed couplings to the third-generation lepton.

The best limit on the coupling strength is obtained for pure-muon coupling scenarios excluding values of $|V_{\mu N}|^2 > 4 \times 10^{-7}$ for Dirac HNLs and $|V_{\mu N}|^2 > 6 \times 10^{-7}$ for Majorana-type HNL particles with $m_N = 10$ GeV at 95% CL. In general, the analysis reached the best sensitivities for HNL particle lifetimes of 1-10 mm and masses

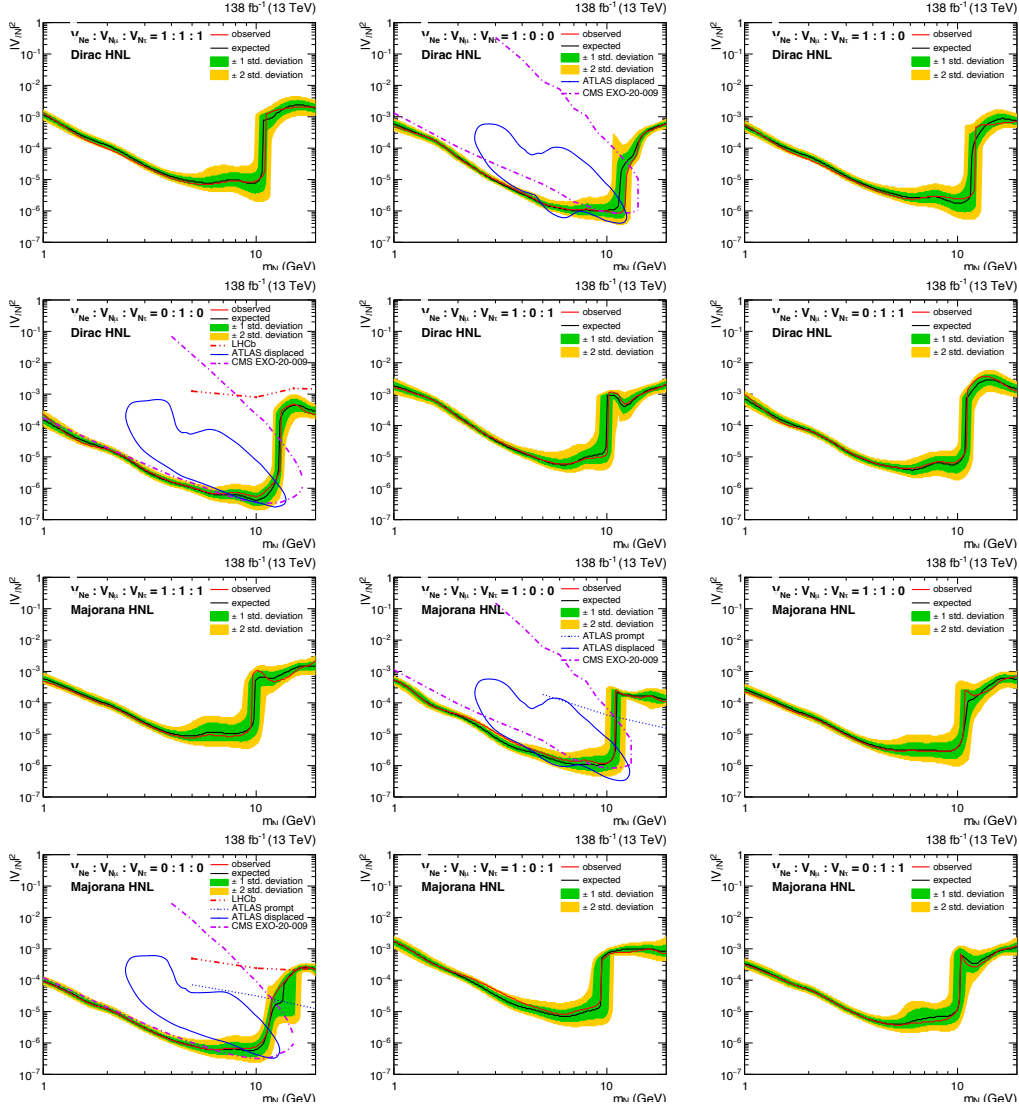


Figure 10.7: Expected and observed limits for Dirac and Majorana HNL for different benchmark coupling scenarios.

between 1-20 GeV. This corresponds to $\sigma_{UL} \lesssim 1\text{pb}$ for $V_e = V_\mu = V_\tau$. Being a tracker-based analysis, the HNL production cross-section limits decrease for lifetimes longer than the tracker size, as well as prompt signals. Higher mass HNL particles typically decay promptly with resolved event topology, whereas lighter HNLs have longer displacements manifesting in merged topologies. The displaced jet tagger has increased performance for the latter models, hence the limits show the best sensitivity up to 10 GeV HNL models, followed by a quick drop towards higher HNL masses leading to prompt scenarios.

Some exclusion limit plots are overlayed with other published results by competing experiments, specifically pure-electron and pure-muon coupling scenarios. No previous search was done on such an extensive set of mixed HNL coupling scenarios, this analysis being the first one to investigate such a broad range of HNL mixing processes.

Exclusion limits are shown on results from the ATLAS experiment [7, 68], LHCb experiment [69] and CMS experiment [6]. The closed loop shape of the ATLAS result is a signature of the search explicitly requiring secondary vertices, hence limiting its sensitivity to prompt HNL decay scenarios. The ATLAS result is the only previous measurement that investigated a limited amount of mixed-coupling HNL scenarios.

Table 10.1 shows the fractional effect of the systematic uncertainties on the HNL signal cross-section expressed via percentage impacts, defined as:

$$\text{percentage impact} = \frac{\Delta\mu^\pm}{\hat{\mu}}, \quad (10.3)$$

where $\Delta\mu^\pm$ is the impact defined in Eq. 5.38 and $\hat{\mu}$ is the nominal value of a parameter of interest when the nuisance parameters are fixed. The largest systematic effects manifest in the ABCD non-closure. However, given the large number of signal channels, the largest overall uncertainty results from low sideband counts in the data-driven background estimation, limiting the analysis statistically. Unclustered energy has a significant impact on the result, as p_T^{miss} variable is a fairly strong input feature governing the performance of the BDT score (see Fig. 8.10a). The JES and JER uncertainties depend on soft jets passing various p_T thresholds, therefore affecting the performance of the tagger score. Displaced lepton identification and reconstruction uncertainties are dominant sources for the signal systematic uncertainty, as the uncertainties on the associated scale factors are large. Detailed impact plots showing all possible systematic uncertainties considered by the analysis are presented in Appendix F.5.

10.3 Summary

This Chapter presents the ABCD data-driven background estimation approach employed to predict the expected background yields in the blinded SR of the analysis. An extensive study of validation tests was performed via closure tests in data in order to gain confidence in the robustness of the background estimation method. After unblinding, results showed consistency with the background-only hypothesis, therefore a number of long-lived Dirac and Majorana-type HNL models were excluded by setting ULs on their production cross-section at 95% CL. Multi-generational HNL coupling scenarios were excluded, besides pure-lepton coupling models. Best exclusion limits were obtained for pure-muon coupling scenarios, and in general, for HNL models with masses around 10 GeV. Observed and expected signal cross-section

Table 10.1: Effect of systematic uncertainties on the HNL signal cross-section expressed as percentage impacts. The impacts on the background yields in the sideband regions are averaged over all 48 categories of the analysis. The bottom five uncertainties affect the background estimation yield, whereas the rest of the entries affect the HNL signal yields.

Systematic uncertainty	Impact (%)		
	2016	2017	2018
Prompt electron reco.	0.0	0.1	0.0
Prompt electron id.	0.1	0.3	0.4
Prompt muon id.	0.0	0.1	0.0
Prompt muon iso.	0.0	0.0	0.0
Trigger	0.1	0.0	0.0
Displaced electron reco.	0.0	0.2	0.3
Displaced electron id.	11.0	0.6	0.6
Displaced muon id.	0.3	0.0	1.7
Displaced muon reco.	0.2	0.3	4.3
Pileup	1.3	3.7	4.9
Jet energy scale	2.5	1.7	4.5
Jet energy resolution	1.0	2.6	1.8
Unclustered energy	3.8	4.8	7.7
Luminosity	1.1	0.6	0.7
PDF	2.8	3.0	1.8
Renormalisation and facotisation scale	0.1	0.3	0.4
Non-closure, boosted	24.8	22.3	14.8
Non-closure, resolved	25.4	24.0	37.6
A (average)	0.8	0.9	1.3
B (average)	4.2	4.8	5.2
C (average)	11.3	8.3	9.2

limits are within agreement, and the analysis shows competitive performance with previously published HNL results at the LHC. Statistical uncertainties show the largest impact on the signal cross-section in ABCD sideband non-closure and limited counts in the data-driven background estimation method, suggesting that the search is statistically limited.

Chapter 11

Conclusions

The Thesis presented a search for long-lived HNL particles in p-p collision data set of 137fb^{-1} integrated luminosity, obtained by the CMS experiment during its Run-2 period (2016-2018). The search covers a wide range of HNL masses, lifetimes, and coupling scenarios, including universal mixing with all SM lepton generations. The final states investigated in the analysis include a prompt lepton the event is triggered on, as well as a displaced lepton and jet. A fine-grained event categorisation scheme is employed to ensure sensitivity to a variety of HNL benchmark models, including Dirac- and Majorana-type particles.

The search takes advantage of a novel DNN displaced jet tagger that was modified to include signatures from displaced leptons in order to accommodate the event topologies of semi-leptonic HNL decays. Domain adaptation is employed to ensure that the DNN tagger output has a low bias towards the simulation set it was trained on and has an accurate response in data, instead.

A BDT-based binary event classifier was developed to be complementary to the jet tagger. While the former aims to separate HNL signal events from the SM background, the latter captures signal jets resulting from the HNL decay. The BDT algorithm was specifically designed to be uncorrelated with the DNN jet tagger, a relevant feature required by the background estimation method employed by the analysis. The performance of the BDT event classifier was evaluated in the signal region of the analysis, showing robust performance, in comparison to a simple cut-based approach or logistic regression. In addition, it was validated in the two control regions of the analysis, providing confidence in the reliability of the method.

The results obtained from the two algorithms were employed in an ABCD data-driven background estimation technique to evaluate the expected yield of signal

and background events in the blinded signal region. No excess of events in the data over the expected background was observed, therefore ULs were set on the HNL production cross-section as a function of the HNL mass and the three coupling strengths to each SM neutrino generation, including both Dirac- and Majorana-type HNL scenarios. Best exclusion limits were obtained for pure-muon coupling scenarios, and in general, for HNL models with masses around 10 GeV. The largest impact on the signal cross-section resulted from statistical uncertainties in the ABCD sideband non-closure and limited counts in the data-driven background estimation method, suggesting that the search was statistically limited. This analysis presents the first search in the CMS collaboration to consider arbitrary coupling combinations involving all three SM lepton generations, corresponding to $\sigma_{\text{UL}} \lesssim 1\text{pb}$ for $V_e = V_\mu = V_\tau$.

A detailed discussion is provided on the treatment of systematic uncertainties in the analysis, with a special focus on those related to displaced lepton identification and reconstruction. Efficiencies and scale factors were derived for both muon and electron identification, via the tag-and-probe method and asymmetric photon conversion events, respectively. Since muons leave “clean” signals throughout the CMS detector with the additional aid of the muon system, their identification scale factors are close to unity throughout the $d_{xy}^{\text{sig}}-p_T$ phase space measured. Electrons, on the other hand, leave complex signatures in the CMS detector due to energy loss via bremsstrahlung radiation, as well as they only reach the ECAL detector system. Therefore, their identification scale factors were slightly poorer, yet still within 10-20% from unity. Muon reconstruction scale factors were derived in the $d_{xy}-p_T$ kinematic phase space, showing good agreement with unity within uncertainties in most bins, except at low- p_T , where track reconstruction algorithms have poorer performance. The scale factors derived for displaced leptons were propagated in the analysis as corrections to signal yields.

In addition, a study on track-to-vertex association algorithms for the Phase-2 upgrade of the CMS Level-1 trigger showed promising improvements compared to the cut-based approach originally considered. The HL-LHC will have pile-up conditions increased by one order of magnitude to the current LHC setup. In order to dramatically reduce the contribution of pile-up tracks, while maintaining competitive physics performance, an end-to-end neural network was proposed for vertex reconstruction and position regression, extended with the track-to-vertex association algorithm. The end-to-end NN was deployed in FPGA firmware, matching stringent latency and memory limits of the Phase-2 Level-1 trigger.

Several possible improvements to the analysis presented in this Thesis could be considered. Firstly, the results are statistically limited, therefore additional data

collection by the Run-3 of the LHC, as well as the HL-LHC would highly benefit the sensitivity to long-lived HNL models. In addition, the potential development of an algorithmically decorrelated BDT from the DNN tagger would result in the possibility of including additional input features that would drastically improve the discrimination power of the event classifier. Finally, a set of reinterpretability studies should be performed on the MVA discriminants employed in the analysis, in order to derive signal efficiencies as a function of HNL lifetime that would aid future HNL model developments by particle physics theorists.

The analysis presented in this Thesis is currently undergoing internal collaboration-wide review within the CMS experiment and is expected to be soon published. The interim results have been released as a CMS physics analysis summary in Ref. [168]. The analysis presented in the paper uses a different background estimation method to that presented in this Thesis, however, the results are very compatible, and the same conclusions may be drawn. Future developments in the field of long-lived HNL research include dedicated triggers for Run-3 of the LHC and the HL-LHC, aiming to enhance the sensitivity to capture its soft decay products. In addition, special data sets such as the B-parking and scouting data collected by the CMS experiment are currently being exploited in long-lived HNL searches. Finally, there are a number of proposed experiments with a physics program dedicated to searches for long-lived particles, including HNLs. Due to the strong theoretical motivation for the existence of HNLs and their vast unexplored phase space, searches for such particles continue to pose a promising topic of research.

Bibliography

- [1] P. Minkowski, “ $\mu \rightarrow e\gamma$ at a Rate of One Out of 10^9 Muon Decays?” *Phys. Lett. B*, vol. 67, pp. 421–428, 1977. DOI: 10.1016/0370-2693(77)90435-X.
- [2] M. Gell-Mann, P. Ramond, and R. Slansky, in *Supergravity: proceedings of the Supergravity Workshop at Stony Brook*, P. Nieuwenhuizen and D. Freedman, Eds., North-Holland, Amsterdam, Netherlands, 1979.
- [3] M. Gell-Mann, P. Ramond, and R. Slansky, *Complex spinors and unified theories*, 2013. arXiv: 1306.4669 [hep-th].
- [4] T. Yanagida, in *Proceedings: Workshop on the Unified Theories and the Baryon Number in the Universe*, O. Sawada and A. Sugamoto, Eds., Tsukuba, Japan: Natl. Lab. High Energy Phys., 1979.
- [5] M. Gronau, C. N. Leung, and J. L. Rosner, “Extending limits on neutral heavy leptons,” *Phys. Rev. D*, vol. 29, pp. 2539–2558, 11 Jun. 1984. DOI: 10.1103/PhysRevD.29.2539.
- [6] CMS Collaboration, “Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{s}=13$ TeV,” *JHEP*, vol. 07, p. 081, 2022. DOI: 10.1007/JHEP07(2022)081. arXiv: 2201.05578 [hep-ex].
- [7] ATLAS Collaboration, “Search for heavy neutral leptons in decays of W bosons produced in 13 TeV pp collisions using prompt and displaced signatures with the ATLAS detector,” *JHEP*, vol. 10, p. 265, 2019. DOI: 10.1007/JHEP10(2019)265. arXiv: 1905.09787 [hep-ex].
- [8] DELPHI Collaboration, “Search for neutral heavy leptons produced in Z decays,” *Z. Phys. C*, vol. 74, pp. 57–71, 1997, [Erratum: *Z.Phys.C* 75, 580 (1997)]. DOI: 10.1007/s002880050370.
- [9] CMS Collaboration, “A deep neural network to search for new long-lived particles decaying to jets,” *Mach. Learn. Sci. Tech.*, vol. 1, p. 035 012, 2020. DOI: 10.1088/2632-2153/ab9023. arXiv: 1912.12238 [hep-ex].
- [10] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013. DOI: 10.1017/CB09781139525367.

- [11] I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics: A Practical Introduction, Volume 1 : From Relativistic Quantum Mechanics to QED, Fourth Edition*. Taylor & Francis, 2013, ISBN: 978-1-4665-1302-0, 978-0-429-18538-0, 978-1-4665-1299-3. DOI: 10.1201/b13717.
- [12] I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics: A Practical Introduction, Volume 2: Non-Abelian Gauge Theories : QCD and The Electroweak Theory, Fourth Edition*. Taylor & Francis, 2013, ISBN: 978-1-4665-1310-5, 978-1-4665-1307-5, 978-0-429-18539-7. DOI: 10.1201/9781466513105.
- [13] J. Schwichtenberg, *Physics from Symmetry* (Undergraduate Lecture Notes in Physics). Cham: Springer International Publishing, 2018, ISBN: 978-3-319-66630-3, 978-3-319-66631-0. DOI: 10.1007/978-3-319-66631-0.
- [14] S. L. Glashow, “The renormalizability of vector meson interactions,” *Nucl. Phys.*, vol. 10, pp. 107–117, 1959. DOI: 10.1016/0029-5582(59)90196-8.
- [15] A. Salam, “Weak and Electromagnetic Interactions,” *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968. DOI: 10.1142/9789812795915_0034.
- [16] S. Weinberg, “A Model of Leptons,” *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, 1967. DOI: 10.1103/PhysRevLett.19.1264.
- [17] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons,” *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., pp. 508–509, 1964. DOI: 10.1103/PhysRevLett.13.508.
- [18] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons,” *Phys. Rev. Lett.*, vol. 13, J. C. Taylor, Ed., pp. 321–323, 1964. DOI: 10.1103/PhysRevLett.13.321.
- [19] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett.*, vol. B716, pp. 30–61, 2012. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].
- [20] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett.*, vol. B716, pp. 1–29, 2012. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].
- [21] K. G. Wilson, “Confinement of Quarks,” *Phys. Rev. D*, vol. 10, J. C. Taylor, Ed., pp. 2445–2459, 1974. DOI: 10.1103/PhysRevD.10.2445.
- [22] LHCb Collaboration, “Observation of New Resonances Decaying to $J/\psi K^+$ and $J/\psi \phi$,” *Phys. Rev. Lett.*, vol. 127, no. 8, p. 082001, 2021. DOI: 10.1103/PhysRevLett.127.082001. arXiv: 2103.01803 [hep-ex].

- [23] LHCb Collaboration, “Observation of a narrow pentaquark state, $P_c(4312)^+$, and of two-peak structure of the $P_c(4450)^+$,” *Phys. Rev. Lett.*, vol. 122, no. 22, p. 222 001, 2019. DOI: 10.1103/PhysRevLett.122.222001. arXiv: 1904.03947 [hep-ex].
- [24] Particle Data Group, “Review of Particle Physics,” *PTEP*, vol. 2020, no. 8, p. 083C01, 2020. DOI: 10.1093/ptep/ptaa104.
- [25] E. Noether, “Invariante Variationsprobleme (in German),” *Nachrichten von der Gesellschaft der Wissenschaften zu Gottingen, Mathematisch-Physikalische Klasse*, 1918. [Online]. Available: <http://eudml.org/doc/59024>.
- [26] G. ’t Hooft and M. J. G. Veltman, “Regularization and Renormalization of Gauge Fields,” *Nucl. Phys. B*, vol. 44, pp. 189–213, 1972. DOI: 10.1016/0550-3213(72)90279-9.
- [27] J. D. Bjorken and E. A. Paschos, “Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon,” *Phys. Rev.*, vol. 185, pp. 1975–1982, 1969. DOI: 10.1103/PhysRev.185.1975.
- [28] C. S. Wu *et al.*, “Experimental Test of Parity Conservation in β Decay,” *Phys. Rev.*, vol. 105, pp. 1413–1414, 1957. DOI: 10.1103/PhysRev.105.1413.
- [29] M. A. Bouchiat and C. Bouchiat, “1. Parity Violation Induced by Weak Neutral Currents in Atomic Physics,” *J. Phys. (France)*, vol. 35, pp. 899–927, 1974. DOI: 10.1051/jphys:019740035012089900.
- [30] ALEPH, DELPHI, L3, OPAL, LEP Electroweak Collaboration, “Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP,” *Phys. Rept.*, vol. 532, pp. 119–244, 2013. DOI: 10.1016/j.physrep.2013.07.004. arXiv: 1302.3415 [hep-ex].
- [31] Working Group on LEP Energy, ALEPH, DELPHI, L3, OPAL Collaboration, “Measurement of the mass of the Z boson and the energy calibration of LEP,” *Phys. Lett. B*, vol. 307, pp. 187–193, 1993. DOI: 10.1016/0370-2693(93)90210-9.
- [32] Particle Data Group, “Review of Particle Physics,” *PTEP*, vol. 2022, p. 083C01, 2022. DOI: 10.1093/ptep/ptac097.
- [33] Muon g-2 Collaboration, “Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm,” *Phys. Rev. Lett.*, vol. 126, no. 14, p. 141 801, 2021. DOI: 10.1103/PhysRevLett.126.141801. arXiv: 2104.03281 [hep-ex].
- [34] Planck Collaboration, “Planck 2018 results - vi. cosmological parameters,” *A&A*, vol. 641, A6, 2020. DOI: 10.1051/0004-6361/201833910. [Online]. Available: <https://doi.org/10.1051/0004-6361/201833910>.

- [35] KATRIN Collaboration, “Direct neutrino-mass measurement with sub-electronvolt sensitivity,” *Nature Phys.*, vol. 18, no. 2, pp. 160–166, 2022. DOI: 10.1038/s41567-021-01463-1. arXiv: 2105.08533 [hep-ex].
- [36] J. A. McCarthy, “Search for double beta decay in Ca^{48} ,” *Phys. Rev.*, vol. 97, pp. 1234–1236, 5 Mar. 1955. DOI: 10.1103/PhysRev.97.1234. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRev.97.1234>.
- [37] V. R. Lazarenko and S. Y. Lukyanov, “An attempt to detect double beta decay in Ca^{48} ,” *JETP*, vol. 22, p. 521, 1966.
- [38] J. J. Gomez-Cadenas *et al.*, “The Search for neutrinoless double beta decay,” *Riv. Nuovo Cim.*, vol. 35, no. 2, pp. 29–98, 2012. DOI: 10.1393/ncr/i2012-10074-9. arXiv: 1109.5515 [hep-ex].
- [39] LSND Collaboration, “Neutrino oscillation results from LSND,” *Nucl. Phys. B Proc. Suppl.*, vol. 91, J. Law, R. W. Ollerhead, and J. J. Simpson, Eds., pp. 198–202, 2001. DOI: 10.1016/S0920-5632(00)00941-5.
- [40] MiniBooNE Collaboration, “Significant Excess of ElectronLike Events in the MiniBooNE Short-Baseline Neutrino Experiment,” *Phys. Rev. Lett.*, vol. 121, no. 22, p. 221 801, 2018. DOI: 10.1103/PhysRevLett.121.221801. arXiv: 1805.12028 [hep-ex].
- [41] MicroBooNE Collaboration, “Search for Neutrino-Induced Neutral-Current Δ Radiative Decay in MicroBooNE and a First Test of the MiniBooNE Low Energy Excess under a Single-Photon Hypothesis,” *Phys. Rev. Lett.*, vol. 128, p. 111 801, 2022. DOI: 10.1103/PhysRevLett.128.111801. arXiv: 2110.00409 [hep-ex].
- [42] M. Goldhaber, L. Grodzins, and A. W. Sunyar, “Helicity of neutrinos,” *Phys. Rev.*, vol. 109, pp. 1015–1017, 3 Feb. 1958. DOI: 10.1103/PhysRev.109.1015. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRev.109.1015>.
- [43] F. T. Avignone, S. R. Elliott, and J. Engel, “Double beta decay, Majorana neutrinos, and neutrino mass,” *Rev. Mod. Phys.*, vol. 80, pp. 481–516, 2 Apr. 2008. DOI: 10.1103/RevModPhys.80.481.
- [44] K. Lundmark, “Über die Bestimmung der Entfernungen, Dimensionen, Massen und Dichtigkeit für die nächstgelegenen anagalaktischen Sternsysteme,” *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, vol. 125, pp. 1–13, Jan. 1930.
- [45] F. Zwicky, “Die Rotverschiebung von extragalaktischen Nebeln,” *Helv. Phys. Acta*, vol. 6, pp. 110–127, 1933. DOI: 10.1007/s10714-008-0707-4.

- [46] F. Zwicky, “On the Masses of Nebulae and of Clusters of Nebulae,” *Astrophys. J.*, vol. 86, pp. 217–246, 1937. DOI: 10.1086/143864.
- [47] V. C. Rubin, N. Thonnard, and W. K. Ford Jr., “Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 ($R = 4\text{ kpc}$) to UGC 2885 ($R = 122\text{ kpc}$),” *Astrophys. J.*, vol. 238, p. 471, 1980. DOI: 10.1086/158003.
- [48] A. N. Taylor *et al.*, “Gravitational lens magnification and the mass of Abell 1689,” *Astrophys. J.*, vol. 501, p. 539, 1998. DOI: 10.1086/305827. arXiv: astro-ph/9801158.
- [49] G. F. Smoot *et al.*, “First results of the COBE satellite measurement of the anisotropy of the cosmic microwave background radiation,” in *2nd Rencontre de Blois on Physical Cosmology: 25th Anniversary of the Cosmic Background Radiation Discovery*, May 1991, pp. 105–120.
- [50] WMAP Collaboration, “Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results,” *Astrophys. J. Suppl.*, vol. 180, pp. 225–245, 2009. DOI: 10.1088/0067-0049/180/2/225. arXiv: 0803.0732 [astro-ph].
- [51] Planck Collaboration, “Planck 2015 results. XIII. Cosmological parameters,” *Astron. Astrophys.*, vol. 594, A13, 2016. DOI: 10.1051/0004-6361/201525830. arXiv: 1502.01589 [astro-ph.CO].
- [52] A. D. Sakharov, “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe,” *Pisma Zh. Eksp. Teor. Fiz.*, vol. 5, pp. 32–35, 1967. DOI: 10.1070/PU1991v034n05ABEH002497.
- [53] J. E. Dodd and B. Gripaios, *The Ideas of Particle Physics*, 4th. Cambridge University Press, Sep. 2020, ISBN: 978-1-108-61627-0, 978-1-108-72740-2. DOI: 10.1017/9781108616270.
- [54] J. H. Christenson *et al.*, “Evidence for the 2π decay of the K_2^0 meson,” *Phys. Rev. Lett.*, vol. 13, pp. 138–140, 4 Jul. 1964. DOI: 10.1103/PhysRevLett.13.138. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.13.138>.
- [55] T2K Collaboration, “Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations,” *Nature*, vol. 580, no. 7803, pp. 339–344, 2020, [Erratum: *Nature* 583, E16 (2020)]. DOI: 10.1038/s41586-020-2177-0. arXiv: 1910.03887 [hep-ex].
- [56] S. Weinberg, “Varieties of Baryon and Lepton Nonconservation,” *Phys. Rev. D*, vol. 22, p. 1694, 1980. DOI: 10.1103/PhysRevD.22.1694.

- [57] J. Schechter and J. W. F. Valle, “Neutrino Masses in $SU(2) \times U(1)$ Theories,” *Phys. Rev. D*, vol. 22, p. 2227, 1980. DOI: 10.1103/PhysRevD.22.2227.
- [58] J. Schechter and J. W. F. Valle, “Neutrino Decay and Spontaneous Violation of Lepton Number,” *Phys. Rev. D*, vol. 25, p. 774, 1982. DOI: 10.1103/PhysRevD.25.774.
- [59] A. M. Abdullahi *et al.*, “The Present and Future Status of Heavy Neutral Leptons,” in *2022 Snowmass Summer Study*, Mar. 2022. arXiv: 2203.08039 [hep-ph].
- [60] T. Asaka and M. Shaposhnikov, “The ν msm, dark matter and baryon asymmetry of the universe,” *Physics Letters B*, vol. 620, no. 1, pp. 17–26, 2005, ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2005.06.020>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0370269305008087>.
- [61] A. Pilaftsis and T. E. J. Underwood, “Resonant leptogenesis,” *Nucl. Phys. B*, vol. 692, pp. 303–345, 2004. DOI: 10.1016/j.nuclphysb.2004.05.029. arXiv: hep-ph/0309342.
- [62] A. Boyarsky, O. Ruchayskiy, and M. Shaposhnikov, “The Role of sterile neutrinos in cosmology and astrophysics,” *Ann. Rev. Nucl. Part. Sci.*, vol. 59, pp. 191–214, 2009. DOI: 10.1146/annurev.nucl.010909.083654. arXiv: 0901.0011 [hep-ph].
- [63] A. Boyarsky *et al.*, “Improved big bang nucleosynthesis constraints on heavy neutral leptons,” *Phys. Rev. D*, vol. 104, no. 2, p. 023517, 2021. DOI: 10.1103/PhysRevD.104.023517. arXiv: 2008.00749 [hep-ph].
- [64] D. Gorbunov and M. Shaposhnikov, “How to find neutral leptons of the ν MSM?” *JHEP*, vol. 10, p. 015, 2007, [Erratum: JHEP 11, 101 (2013)]. DOI: 10.1088/1126-6708/2007/10/015. arXiv: 0705.1729 [hep-ph].
- [65] L. Lee *et al.*, “Collider Searches for Long-Lived Particles Beyond the Standard Model,” *Prog. Part. Nucl. Phys.*, vol. 106, pp. 210–255, 2019. DOI: 10.1016/j.pnpnp.2019.02.006. arXiv: 1810.12602 [hep-ph].
- [66] *An overview of detector signals not predicted by the standard model*. Nov. 2022. [Online]. Available: <https://www.oead.ac.at/en/hephy/research/cms-experiment-at-the-lhc/data-analysis-group>.
- [67] A. Abada *et al.*, “Inclusive displaced vertex searches for heavy neutral leptons at the LHC,” *Journal of High Energy Physics*, vol. 2019, no. 1, p. 93, 2019.

- [68] ATLAS Collaboration, “Search for heavy neutral leptons in decays of W bosons using a dilepton displaced vertex in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector,” Apr. 2022. arXiv: 2204.11988 [hep-ex].
- [69] LHCb Collaboration, “Search for heavy neutral leptons in $W^+ \rightarrow \mu^+ \mu^\pm \text{jet}$ decays,” *Eur. Phys. J. C*, vol. 81, no. 3, p. 248, 2021. DOI: 10.1140/epjc/s10052-021-08973-5. arXiv: 2011.05263 [hep-ex].
- [70] “LHC Machine,” *JINST*, vol. 3, L. Evans and P. Bryant, Eds., S08001, 2008. DOI: 10.1088/1748-0221/3/08/S08001.
- [71] S. Fartoukh *et al.*, “LHC Configuration and Operational Scenario for Run 3,” CERN, Geneva, Tech. Rep., 2021. [Online]. Available: <https://cds.cern.ch/record/2790409>.
- [72] H. Xie, Y. Song, and S. Wang, “Design evolution and analysis of the iter cryostat support system,” *Plasma Science and Technology*, vol. 17, no. 12, p. 1061, Dec. 2015. DOI: 10.1088/1009-0630/17/12/14. [Online]. Available: <https://dx.doi.org/10.1088/1009-0630/17/12/14>.
- [73] S. Griffith and B. Doshi, *The largest fridge around*, Apr. 2023. [Online]. Available: <https://www.iter.org/newsline/189/858>.
- [74] *Public CMS Luminosity Information*, Dec. 2022. [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [75] CMS Collaboration, “The CMS Experiment at the CERN LHC,” *JINST*, vol. 3, S08004, 2008. DOI: 10.1088/1748-0221/3/08/S08004.
- [76] CMS Collaboration, “Technical Proposal for the Phase-II Upgrade of the CMS Detector,” D. Contardo *et al.*, Eds., Jun. 2015. DOI: 10.17181/CERN.VU8I.D59J.
- [77] CMS Collaboration, “CMS Physics: Technical design report volume 1: Detector performance and software,” Technical Design Report CMS, 2006. [Online]. Available: <https://cds.cern.ch/record/922757>.
- [78] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector,” *JINST*, vol. 12, no. 10, P10003, 2017. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].
- [79] CMS Collaboration, “CMS, the magnet project: Technical design report,” May 1997.
- [80] CMS Collaboration, “The CMS tracker system project: Technical Design Report,” M. Mannelli *et al.*, Eds., 1997.
- [81] *Sketches of the CMS Tracker Detector*, Jan. 2023. [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK>.

- [82] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker,” *JINST*, vol. 9, no. 10, P10009, 2014. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].
- [83] CMS Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report,” 1997.
- [84] CMS Collaboration, “The CMS ECAL performance with examples,” *JINST*, vol. 9, P. Govoni *et al.*, Eds., p. C02008, 2014. DOI: 10.1088/1748-0221/9/02/C02008.
- [85] P. Adzic *et al.*, “Energy resolution of the barrel of the CMS electromagnetic calorimeter,” *JINST*, vol. 2, P04004, 2007. DOI: 10.1088/1748-0221/2/04/P04004.
- [86] CMS Collaboration, “The CMS hadron calorimeter project: Technical Design Report,” 1997.
- [87] CMS Collaboration, “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data,” *JINST*, vol. 5, T03012, 2010. DOI: 10.1088/1748-0221/5/03/T03012. arXiv: 0911.4991 [physics.ins-det].
- [88] CMS ECAL Collaboration, “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c,” *J. Phys. Conf. Ser.*, vol. 160, M. Frater-nali, G. Gaudio, and M. Livan, Eds., p. 012 056, 2009. DOI: 10.1088/1742-6596/160/1/012056.
- [89] CMS Collaboration, “The CMS muon project: Technical Design Report,” 1997.
- [90] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV,” *JINST*, vol. 13, no. 06, P06015, 2018. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].
- [91] CMS Collaboration, “The CMS trigger system,” *JINST*, vol. 12, no. 01, P01020, 2017. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].
- [92] CMS Collaboration, “CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems,” Dec. 2000.
- [93] CMS Collaboration, “CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger,” P. Sphicas, Ed., Dec. 2002.
- [94] CMS Collaboration, “The CMS Level-1 Trigger for LHC Run II,” *PoS*, vol. ICHEP2016, p. 242, 2016. DOI: 10.22323/1.282.0242.

- [95] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV,” *JINST*, vol. 15, no. 10, P10017, 2020. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].
- [96] CMS Collaboration, “CMS Technical Design Report for the Level-1 Trigger Upgrade,” no. CERN-LHCC-2013-011. CMS-TDR-12, A. Tapper, Ed., Jun. 2013. [Online]. Available: <https://cds.cern.ch/record/1556311>.
- [97] A. Afaq *et al.*, “The CMS High Level Trigger System,” *IEEE Trans. Nucl. Sci.*, vol. 55, pp. 172–175, 2008. DOI: 10.1109/TNS.2007.910980.
- [98] I. Zurbano Fernandez *et al.*, “High-Luminosity Large Hadron Collider (HL-LHC): Technical design report,” vol. 10/2020, I. Béjar Alonso *et al.*, Eds., Dec. 2020. DOI: 10.23731/CYRM-2020-0010.
- [99] CMS Collaboration, “The Phase-2 Upgrade of the CMS Level-1 Trigger,” CERN, Geneva, Tech. Rep., 2020, Final version. [Online]. Available: <https://cds.cern.ch/record/2714892>.
- [100] S. P. Summers, “Application of FPGAs to triggering in high energy physics,” Ph.D. dissertation, Imperial Coll., London, 2018. DOI: 10.25560/66689.
- [101] E. D. Sontag, “Rudolf E. Kalman and His Students [Historical Perspectives],” *IEEE Control Systems Magazine*, vol. 30, no. 2, pp. 87–88, 2010. DOI: 10.1109/MCS.2010.935885.
- [102] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting,” *Nucl. Instrum. Meth. A*, vol. 262, pp. 444–450, 1987. DOI: 10.1016/0168-9002(87)90887-4.
- [103] P. Billoir, “Progressive track recognition with a Kalman like fitting procedure,” *Comput. Phys. Commun.*, vol. 57, pp. 390–394, 1989. DOI: 10.1016/0010-4655(89)90249-X.
- [104] CMS Collaboration, “CMS track reconstruction performance during Run 2 and developments for Run 3,” *PoS*, vol. ICHEP2020, p. 733, 2021. DOI: 10.22323/1.390.0733. arXiv: 2012.07035 [physics.ins-det].
- [105] W. Adam *et al.*, “Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC,” *eConf*, vol. C0303241, TULT009, 2003. DOI: 10.1088/0954-3899/31/9/N01. arXiv: physics/0306087.
- [106] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems,” *IEEE Proc.*, vol. 86, no. 11, pp. 2210–2239, 1998. DOI: 10.1109/5.726788.
- [107] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, “Adaptive vertex fitting,” *J. Phys. G*, vol. 34, N343, 2007. DOI: 10.1088/0954-3899/34/12/N01.

- [108] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC,” *JINST*, vol. 16, no. 05, P05014, 2021. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].
- [109] V. Čepaitis, “Flavour-universal search for heavy neutral leptons with a deep neural network-based displaced jet tagger with the CMS experiment,” Ph.D. dissertation, Imperial Coll., London, 2021. DOI: 10.25560/99887.
- [110] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_T jet clustering algorithm,” *JHEP*, vol. 04, p. 063, 2008. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].
- [111] CMS Collaboration, “Pileup removal algorithms,” Tech. Rep. CMS-PAS-JME-14-001, 2014, <https://cds.cern.ch/record/1751454>. [Online]. Available: <https://cds.cern.ch/record/1751454>.
- [112] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV,” *JINST*, vol. 12, no. 02, P02014, 2017. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].
- [113] CMS Collaboration, “Recommended Jet Energy Corrections and Uncertainties For Data and MC,” [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- [114] CMS Collaboration, “Jet energy scale and resolution performances with 13 TeV data,” Jun. 2016. [Online]. Available: <http://cds.cern.ch/record/2160347>.
- [115] CMS Collaboration, “Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018,” 2020. [Online]. Available: <https://cds.cern.ch/record/2715872>.
- [116] E. Bols *et al.*, “Jet Flavour Classification Using DeepJet,” *JINST*, vol. 15, no. 12, P12012, 2020. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex].
- [117] CMS Collaboration, “Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector,” *JINST*, vol. 14, no. 07, P07004, 2019. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].
- [118] A. Buckley *et al.*, “General-purpose event generators for LHC physics,” *Phys. Rept.*, vol. 504, pp. 145–233, 2011. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].

- [119] M. Herquet and F. Maltoni, “MadGraph/MadEvent : A multipurpose event generator,” *Nucl. Phys. B Proc. Suppl.*, vol. 179-180, D. d’Enterria, M. Klasen, and K. Piotrkowski, Eds., pp. 211–217, 2008. DOI: 10.1016/j.nuclphysbps.2008.07.026.
- [120] B. Andersson *et al.*, “Parton Fragmentation and String Dynamics,” *Phys. Rept.*, vol. 97, pp. 31–145, 1983. DOI: 10.1016/0370-1573(83)90080-7.
- [121] J. Alwall *et al.*, “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions,” *Eur. Phys. J. C*, vol. 53, pp. 473–500, 2008. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph].
- [122] K. Danziger, S. Höche, and F. Siegert, “Reducing negative weights in Monte Carlo event generation with Sherpa,” Oct. 2021. arXiv: 2110.15211 [hep-ph].
- [123] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements,” *Eur. Phys. J. C*, vol. 76, no. 3, p. 155, 2016. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].
- [124] T. Sjöstrand *et al.*, “An introduction to PYTHIA 8.2,” *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].
- [125] NNPDF Collaboration, “Parton distributions for the LHC Run II,” *JHEP*, vol. 04, p. 040, 2015. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].
- [126] NNPDF Collaboration, “Parton distributions from high-precision collider data,” *Eur. Phys. J. C*, vol. 77, no. 10, p. 663, 2017. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].
- [127] GEANT4 Collaboration, “GEANT4—a simulation toolkit,” *Nucl. Instrum. Meth. A*, vol. 506, pp. 250–303, 2003. DOI: 10.1016/S0168-9002(03)01368-8.
- [128] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition* (Springer Series in Statistics). Springer New York, 2009, ISBN: 9780387848587.
- [129] Y. Freund and R. E. Schapire, “Experiments with a new boosting algorithm,” in *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*, ser. ICML’96, Bari, Italy: Morgan Kaufmann Publishers Inc., 1996, pp. 148–156, ISBN: 1558604197.

- [130] T. Chen and C. Guestrin, "Xgboost," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016. DOI: 10.1145/2939672.2939785. [Online]. Available: <http://dx.doi.org/10.1145/2939672.2939785>.
- [131] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943, ISSN: 1522-9602. DOI: 10.1007/BF02478259. [Online]. Available: <https://doi.org/10.1007/BF02478259>.
- [132] C. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Springer New York, 2016, ISBN: 9781493938438.
- [133] M. Ester *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. of 2nd International Conference on Knowledge Discovery and*, 1996, pp. 226–231.
- [134] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, Jan. 2016, ISSN: 1532-4435. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2946645.2946704>.
- [135] "Pearson's correlation coefficient," in *Encyclopedia of Public Health*, W. Kirch, Ed. Dordrecht: Springer Netherlands, 2008, pp. 1090–1091, ISBN: 978-1-4020-5614-7. DOI: 10.1007/978-1-4020-5614-7_2569. [Online]. Available: https://doi.org/10.1007/978-1-4020-5614-7_2569.
- [136] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.
- [137] D. W. Scott, *Multivariate Density Estimation. Theory, Practice, and Visualization*. Wiley, 1992.
- [138] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, vol. 26, no. 4, pp. 404–413, Dec. 1934, ISSN: 0006-3444. DOI: 10.1093/biomet/26.4.404. eprint: <https://academic.oup.com/biomet/article-pdf/26/4/404/823407/26-4-404.pdf>. [Online]. Available: <https://doi.org/10.1093/biomet/26.4.404>.
- [139] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 2* (Wiley Series in Probability and Statistics). Wiley, 1995, ISBN: 9780471584940. [Online]. Available: <https://books.google.fr/books?id=BTANEAAAQBAJ>.
- [140] G. Cowan, *Statistical Data Analysis* (Oxford science publications). Clarendon Press, 1998, ISBN: 9780198501558.

- [141] L. Lyons, *Statistics for Nuclear and Particle Physicists*. Cambridge University Press, 1989, ISBN: 9781316098578.
- [142] J. Neyman and E. S. Pearson, “On the Problem of the Most Efficient Tests of Statistical Hypotheses,” *Phil. Trans. Roy. Soc. Lond. A*, vol. 231, no. 694-706, pp. 289–337, 1933. DOI: 10.1098/rsta.1933.0009.
- [143] S. Choi and H. Oh, “Improved Extrapolation Methods of Data-driven Background Estimation in High-Energy Physics,” Jun. 2019. arXiv: 1906.10831 [hep-ph].
- [144] L. Demortier, “P values and nuisance parameters,” in *Statistical issues for LHC physics. Proceedings, Workshop, PHYSTAT-LHC, Geneva, Switzerland, June 27-29, 2007*, 2008, p. 23. DOI: 10.5170/CERN-2008-001. [Online]. Available: <http://cds.cern.ch/record/1099967/files/p23.pdf>.
- [145] G. Cowan *et al.*, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C*, vol. 71, p. 1554, 2011, [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)]. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].
- [146] S. Baker and R. D. Cousins, “Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms,” *Nucl. Instrum. Meth.*, vol. 221, pp. 437–442, 1984. DOI: 10.1016/0167-5087(84)90016-4.
- [147] G. J. Feldman and R. D. Cousins, “A Unified approach to the classical statistical analysis of small signals,” *Phys. Rev. D*, vol. 57, pp. 3873–3889, 1998. DOI: 10.1103/PhysRevD.57.3873. arXiv: physics/9711021.
- [148] A. Wald, “Tests of statistical hypotheses concerning several parameters when the number of observations is large,” *Transactions of the American Mathematical Society*, vol. 54, no. 3, pp. 426–482, 1943, ISSN: 00029947. [Online]. Available: <http://www.jstor.org/stable/1990256>.
- [149] G. Cowan, *Discovery sensitivity for a counting experiment with background uncertainty*, May 2012. [Online]. Available: <https://www.pp.rhul.ac.uk/~cowan/stat/medsig/medsigNote.pdf>.
- [150] I. Tomalin *et al.*, “An FPGA based track finder for the L1 trigger of the CMS experiment at the High Luminosity LHC,” *JINST*, vol. 12, P12019, 2017. DOI: 10.1088/1748-0221/12/12/P12019.
- [151] A. Shtipliyski, “Systems and algorithms for low-latency event reconstruction for upgrades of the Level-1 trigger of the CMS experiment at CERN,” Presented 2019, 2019. [Online]. Available: <https://cds.cern.ch/record/2708430>.

- [152] CMS Collaboration, “The Phase-2 Upgrade of the CMS L1 Trigger Interim Technical Design Report,” CERN, Geneva, Tech. Rep., 2017, This is the CMS Interim TDR devoted to the upgrade of the CMS L1 trigger in view of the HL-LHC running, as approved by the LHCC. DOI: 10.17181/CERN.UUWZ.FMIS. [Online]. Available: <https://cds.cern.ch/record/2283192>.
- [153] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *J. Machine Learning Res.*, vol. 12, pp. 2825–2830, 2011. arXiv: 1201.0490 [cs.LG].
- [154] CMS Collaboration, “Neural Network-Based Primary Vertex Reconstruction with FPGAs for the Upgrade of the CMS Level-1 Trigger System,” *J. Phys. Conf. Ser.*, vol. 2438, no. 1, p. 012 106, 2023. DOI: 10.1088/1742-6596/2438/1/012106.
- [155] R. Ruiz, “HeavyN,” [Online]. Available: <http://feynrules.irmp.ucl.ac.be/wiki/HeavyN>.
- [156] O. Mattelaer, “On the maximal use of Monte Carlo samples: re-weighting events at NLO accuracy,” *Eur. Phys. J. C*, vol. 76, no. 12, p. 674, 2016. DOI: 10.1140/epjc/s10052-016-4533-7. arXiv: 1607.00763 [hep-ph].
- [157] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO,” *JHEP*, vol. 12, p. 061, 2012. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph].
- [158] J. F. Donoghue, F. E. Low, and S.-Y. Pi, “Tensor Analysis of Hadronic Jets in Quantum Chromodynamics,” *Phys. Rev. D*, vol. 20, p. 2759, 1979. DOI: 10.1103/PhysRevD.20.2759.
- [159] ATLAS Collaboration, “Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at $\sqrt{s} = 7$ TeV,” *Eur. Phys. J. C*, vol. 72, p. 2211, 2012. DOI: 10.1140/epjc/s10052-012-2211-y. arXiv: 1206.2135 [hep-ex].
- [160] E. Schlieckau, “First Candidates of Top Quark Events in the Fully Hadronic Decay Channel at the LHC,” Ph.D. dissertation, Hamburg U., 2010.
- [161] CMS Collaboration, “Performance of missing transverse momentum reconstruction in events containing a photon and jets collected by CMS during proton-proton collisions at $\sqrt{s} = 13$ TeV in 2018,” Jul. 2020. [Online]. Available: <https://cds.cern.ch/record/2723010>.
- [162] P. Baldi *et al.*, “Parameterized neural networks for high-energy physics,” *Eur. Phys. J. C*, vol. 76, no. 5, p. 235, 2016. DOI: 10.1140/epjc/s10052-016-4099-4. arXiv: 1601.07913 [hep-ex].

- [163] CMS Collaboration, “Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS,” *Eur. Phys. J. C*, vol. 81, no. 9, p. 800, 2021. DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927 [hep-ex].
- [164] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV,” CERN, Geneva, Tech. Rep., 2018. [Online]. Available: <https://cds.cern.ch/record/2621960>.
- [165] CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV,” CERN, Geneva, Tech. Rep., 2019. [Online]. Available: <https://cds.cern.ch/record/2676164>.
- [166] *J/ψ mass peak fit with IMINUIT - fit parameter setup*, May 2022. [Online]. Available: <https://github.com/mieskolainen/icenet/blob/master/configs/peakfit/tune0.yml>.
- [167] *J/ψ mass peak fit with IMINUIT - algorithm*, May 2022. [Online]. Available: <https://github.com/mieskolainen/icenet/blob/master/icefit/peakfit.py>.
- [168] CMS Collaboration, “Search for long-lived heavy neutral leptons with lepton flavour conserving or violating decays to a jet and an electron, muon, or tau lepton,” CERN, Geneva, Tech. Rep., 2023. [Online]. Available: <https://cds.cern.ch/record/2852843>.

Appendix A

Simple error propagation equation

Error propagation equation for function $f(x, y)$ where x and y are two uncorrelated variables:

$$\sigma(f(x, y)) = \pm \sqrt{\left(\frac{\partial f}{\partial x} \sigma(x)\right)^2 + \left(\frac{\partial f}{\partial y} \sigma(y)\right)^2} \quad (\text{A.1})$$

For the case of $f(x, y) = \frac{x}{y}$, the error on the ratio becomes:

$$\sigma(f(x, y)) = \frac{x}{y} \sqrt{\left(\frac{\sigma(x)}{x}\right)^2 + \left(\frac{\sigma(y)}{y}\right)^2} \quad (\text{A.2})$$

Given $x = \varepsilon_{\text{data}}$ and $y = \varepsilon_{\text{MC}}$, the uncertainty on the scale factor, which is their ratio, becomes:

$$\sigma(f(\varepsilon_{\text{data}}, \varepsilon_{\text{MC}})) = \frac{\varepsilon_{\text{data}}}{\varepsilon_{\text{MC}}} \sqrt{\left(\frac{\sigma(\varepsilon_{\text{data}})}{\varepsilon_{\text{data}}}\right)^2 + \left(\frac{\sigma(\varepsilon_{\text{MC}})}{\varepsilon_{\text{MC}}}\right)^2} \quad (\text{A.3})$$

The error on the MC efficiency is constant for up and down variation, whereas the Clopper-Pearson binomial error is asymmetric for the efficiency on data. The uncertainty on the scale factor is evaluated separately for the up and down variation.

Appendix B

Heatmaps for the track-to-vertex association studies

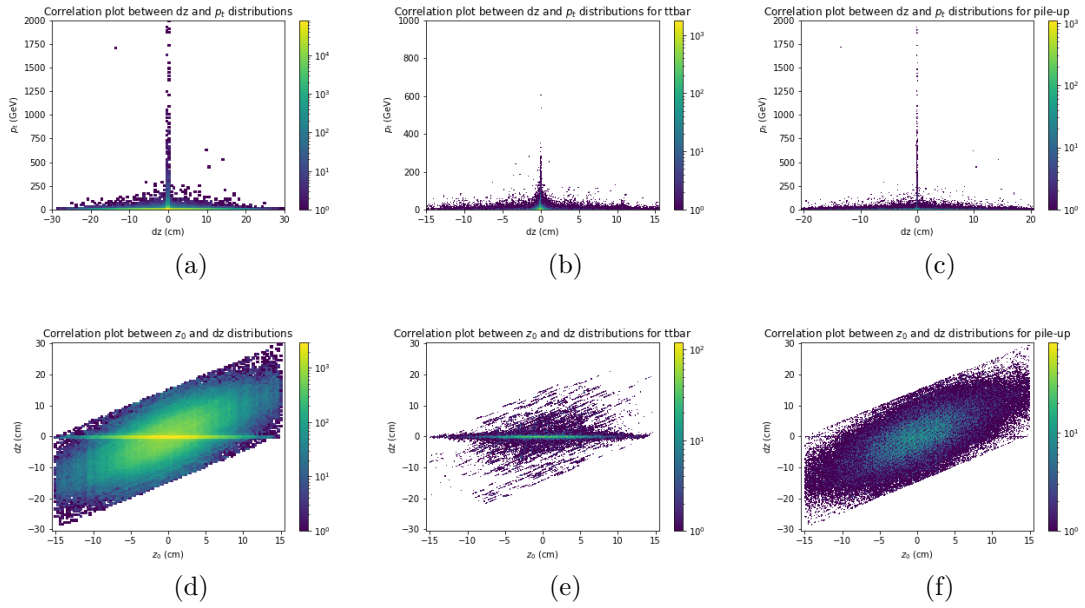


Figure B.1: Plots presenting the 2D distribution and lack of correlation between δz and p_T , as well as z_0 and δz for $t\bar{t}$ tracks and PU. There is a tendency towards $\delta z = 0$, particularly for some of the $t\bar{t}$ tracks.

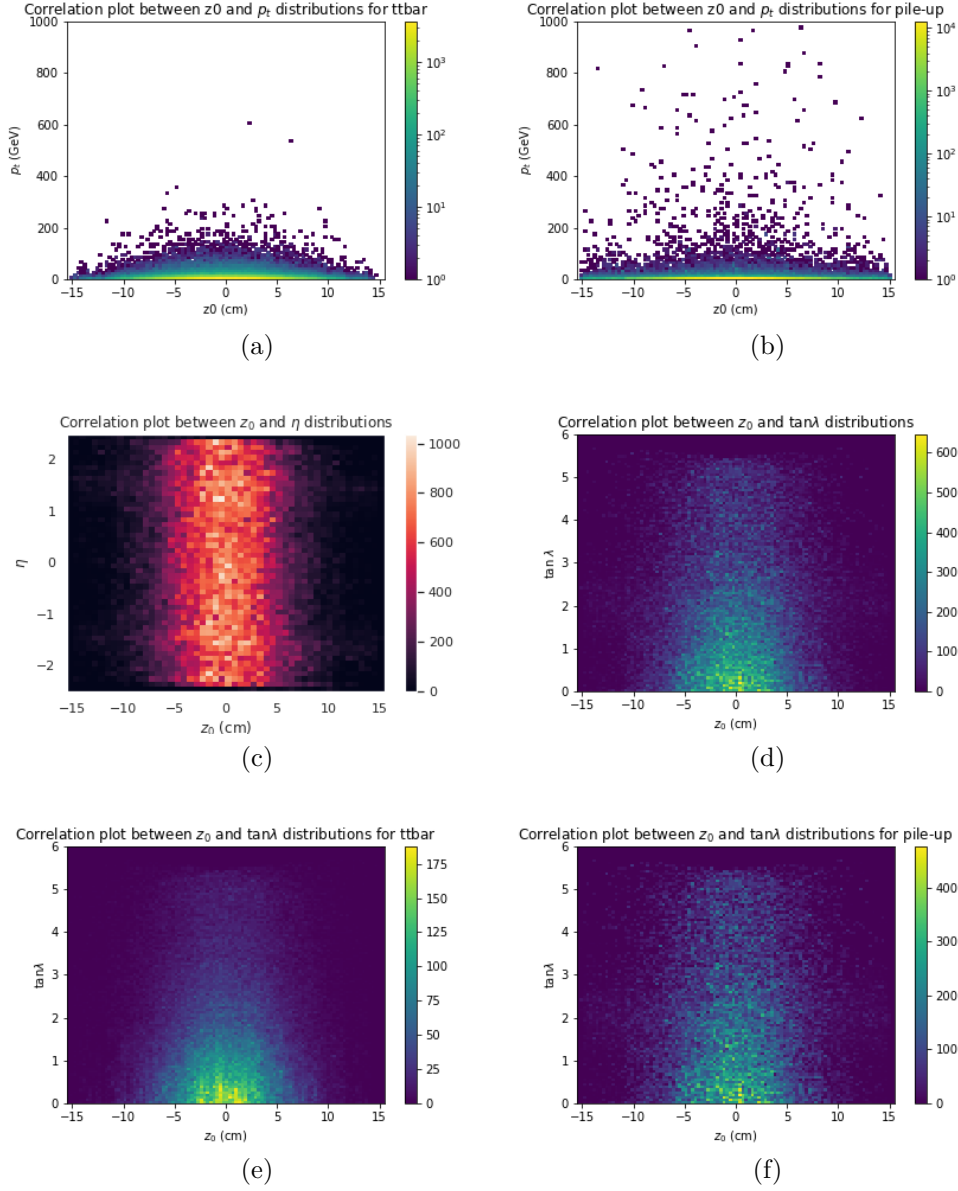


Figure B.2: Heatmaps presenting the uniform distributions of the p_T vs z_0 , η vs z_0 , as well as $\tan\lambda$ vs z_0 proving no correlation between them for $t\bar{t}$ tracks and PU.

Appendix C

Control region validation plots

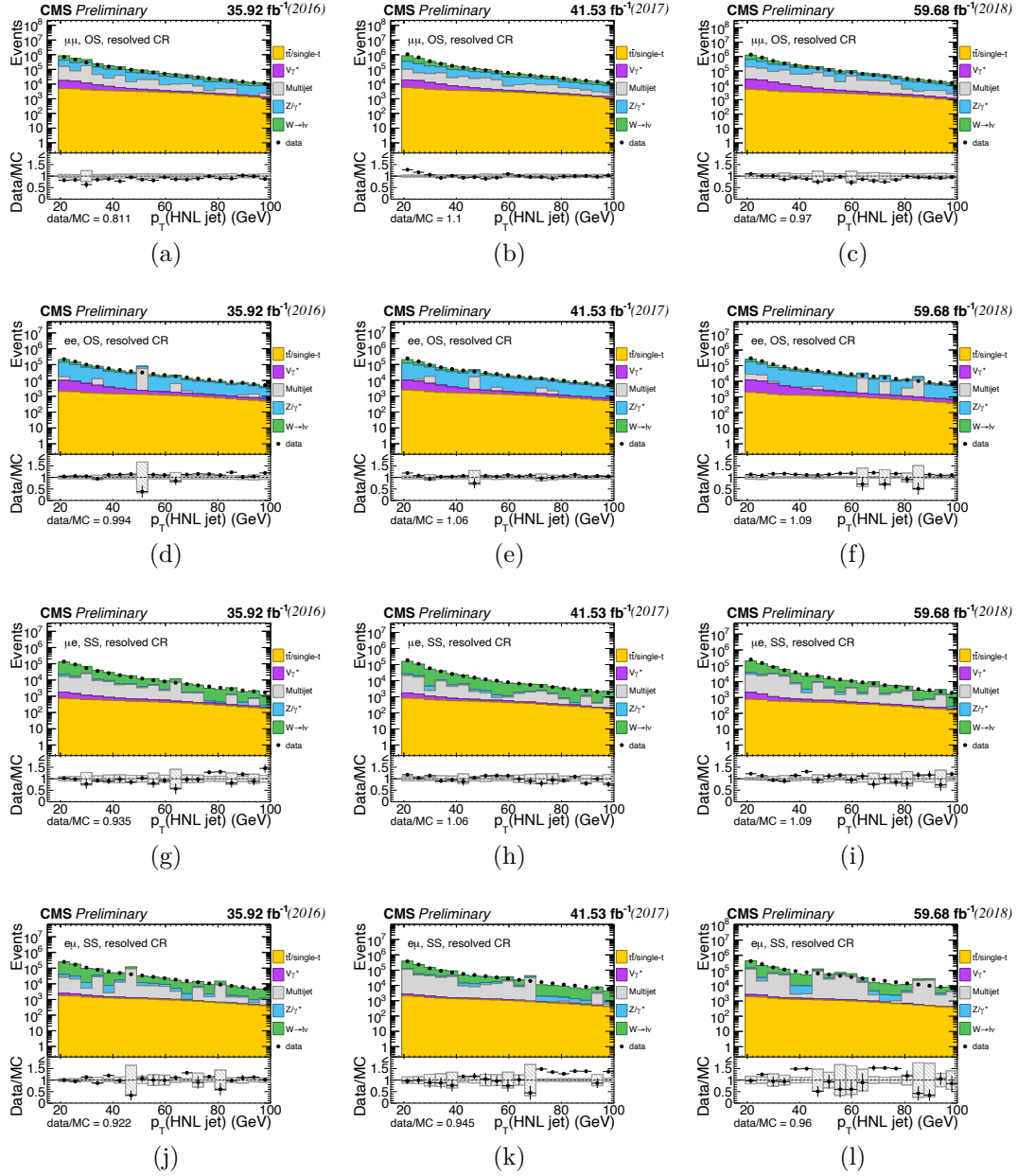


Figure C.1: Distribution of the HNL candidate jet transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

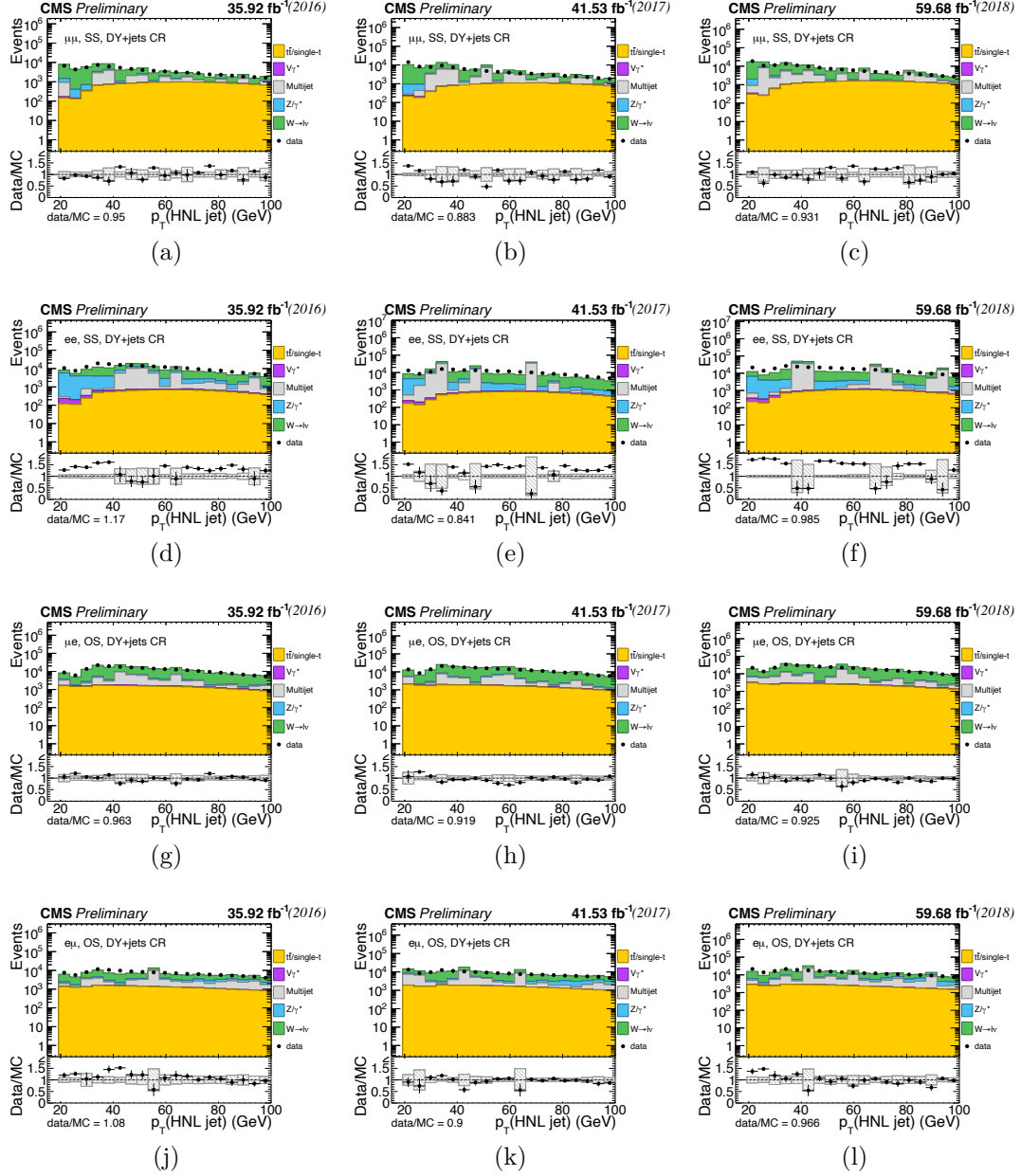


Figure C.2: Distribution of the HNL candidate jet transverse momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.

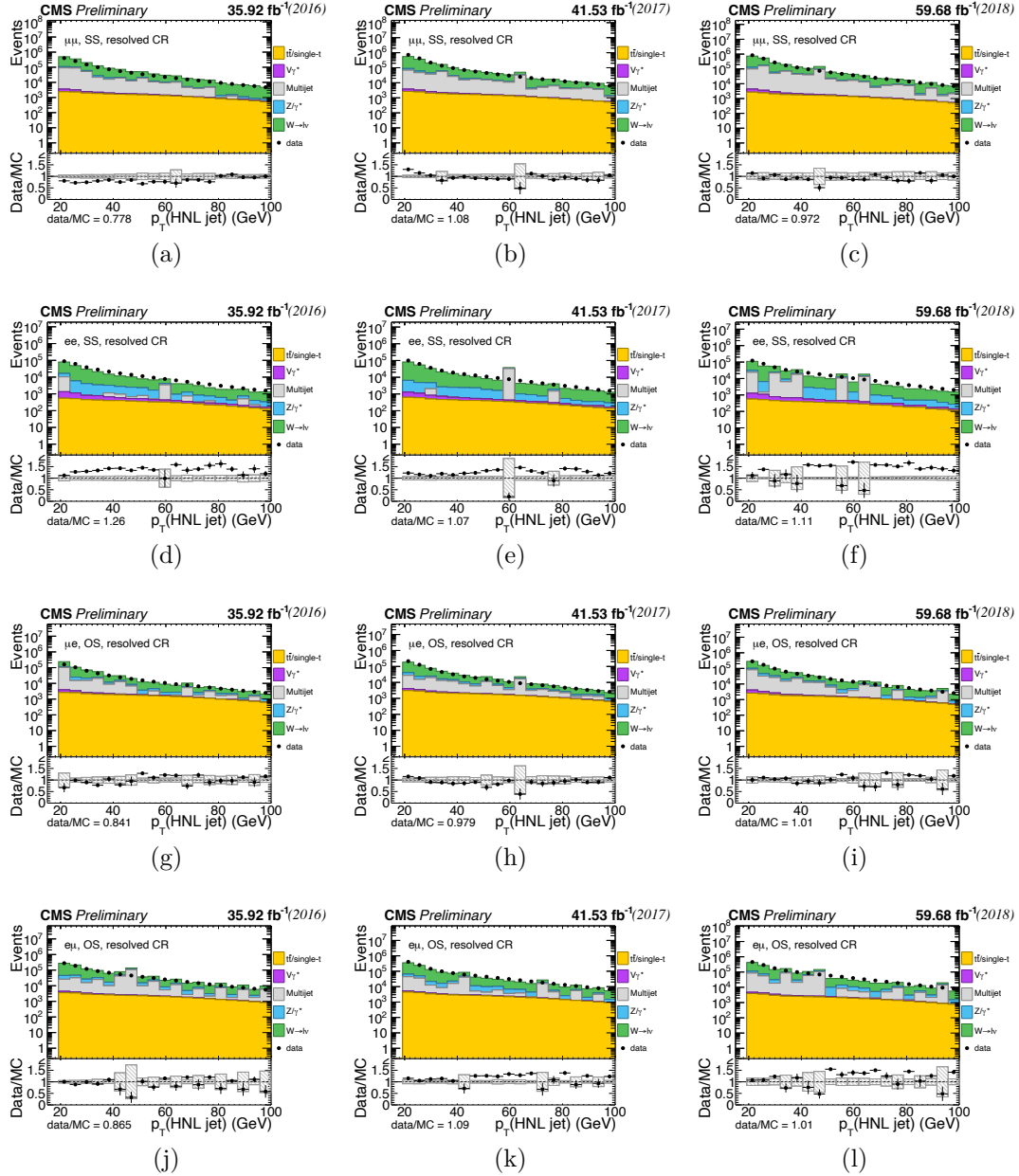


Figure C.3: Distribution of the HNL candidate jet transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.

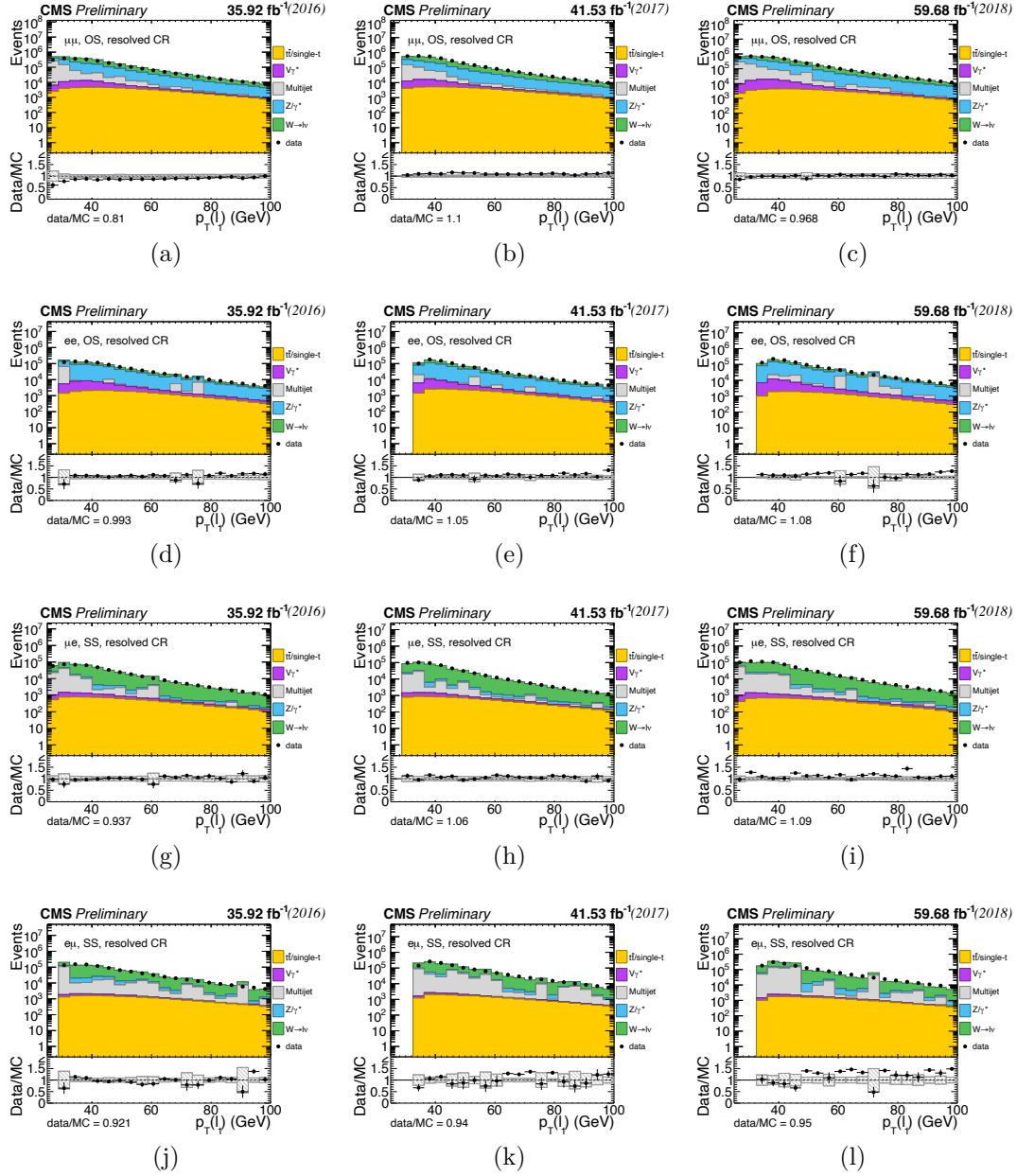


Figure C.4: Distribution of the ℓ_1 transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

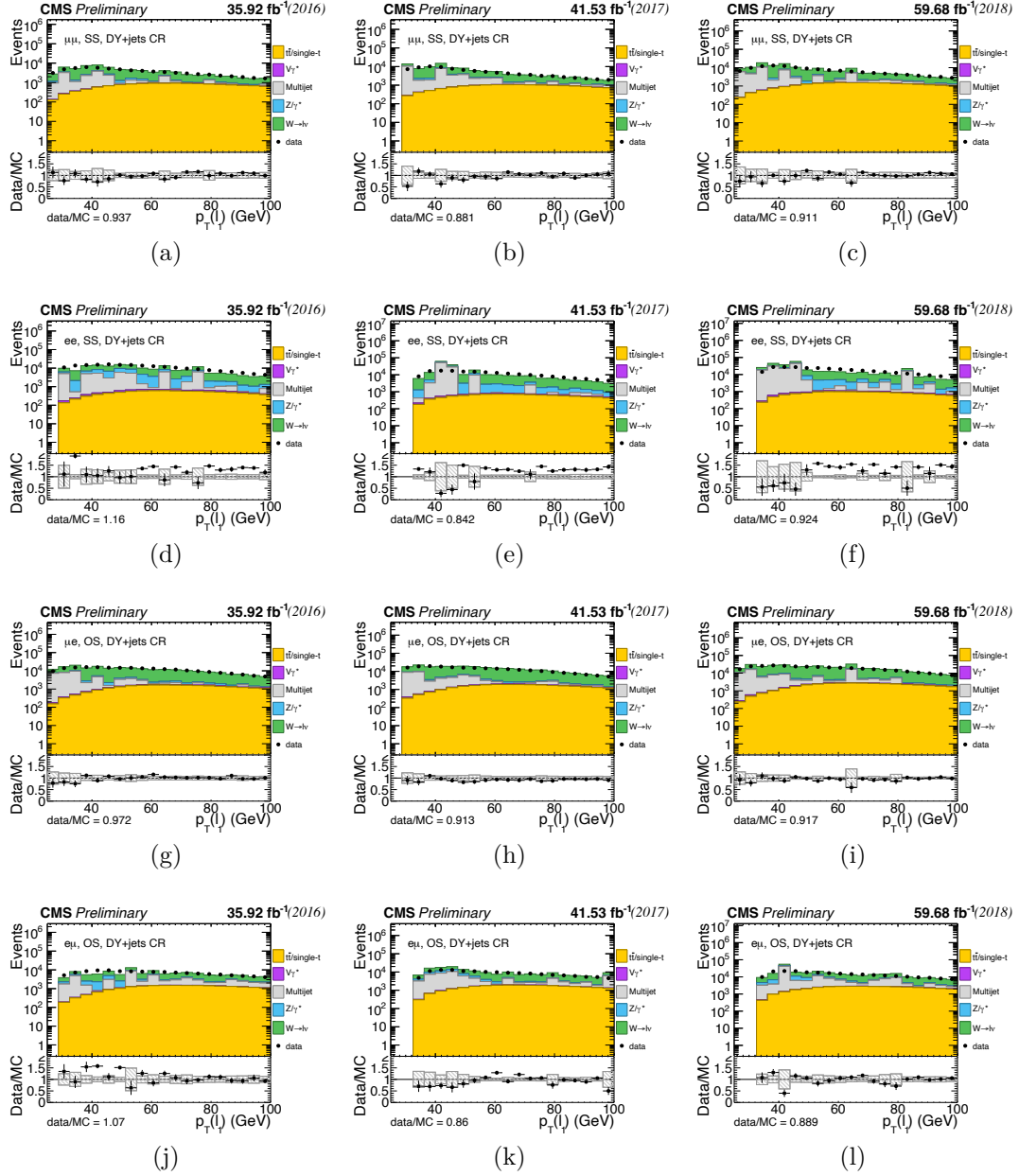


Figure C.5: Distribution of the ℓ_1 momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.

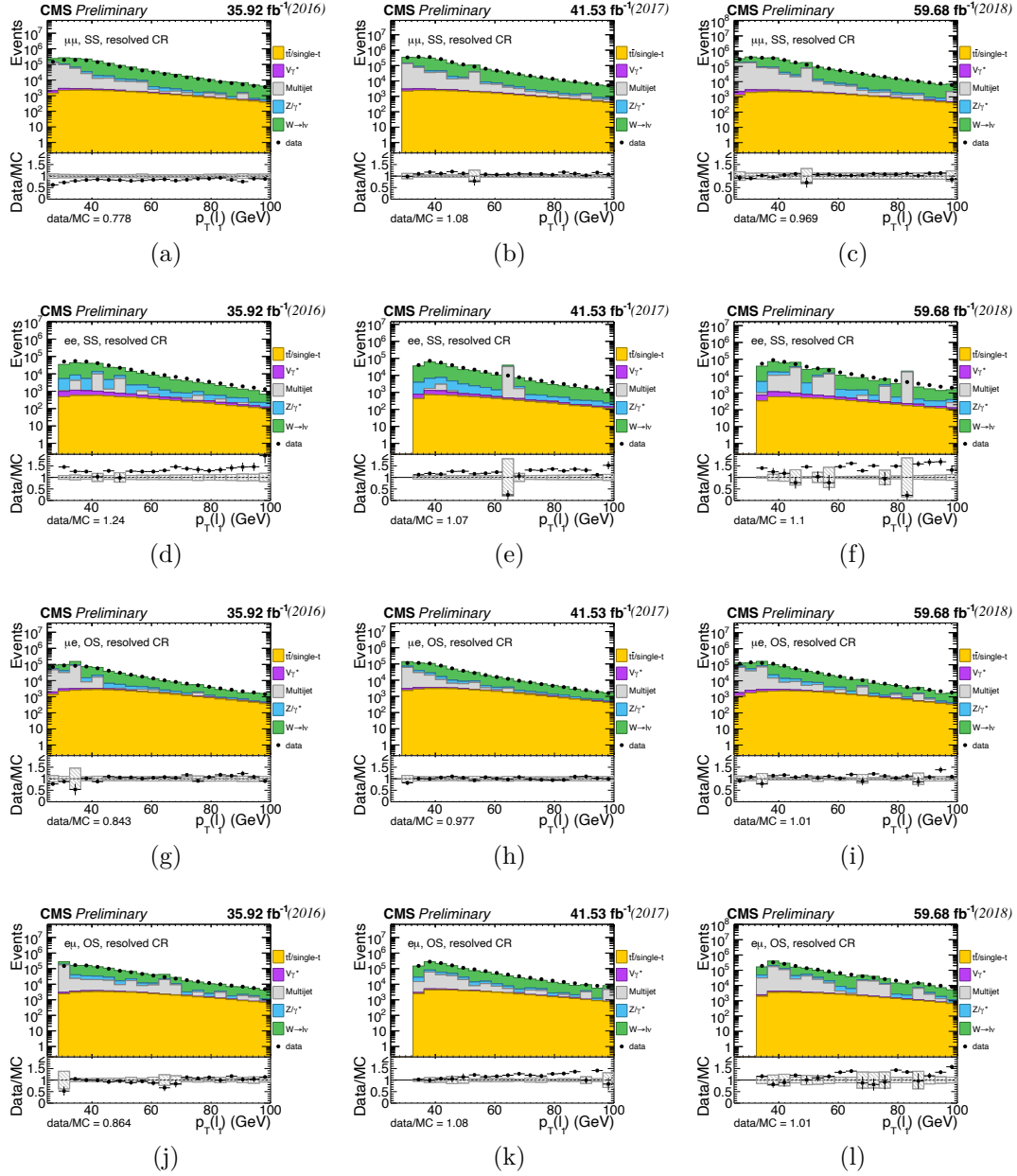


Figure C.6: Distribution of the ℓ_1 transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.

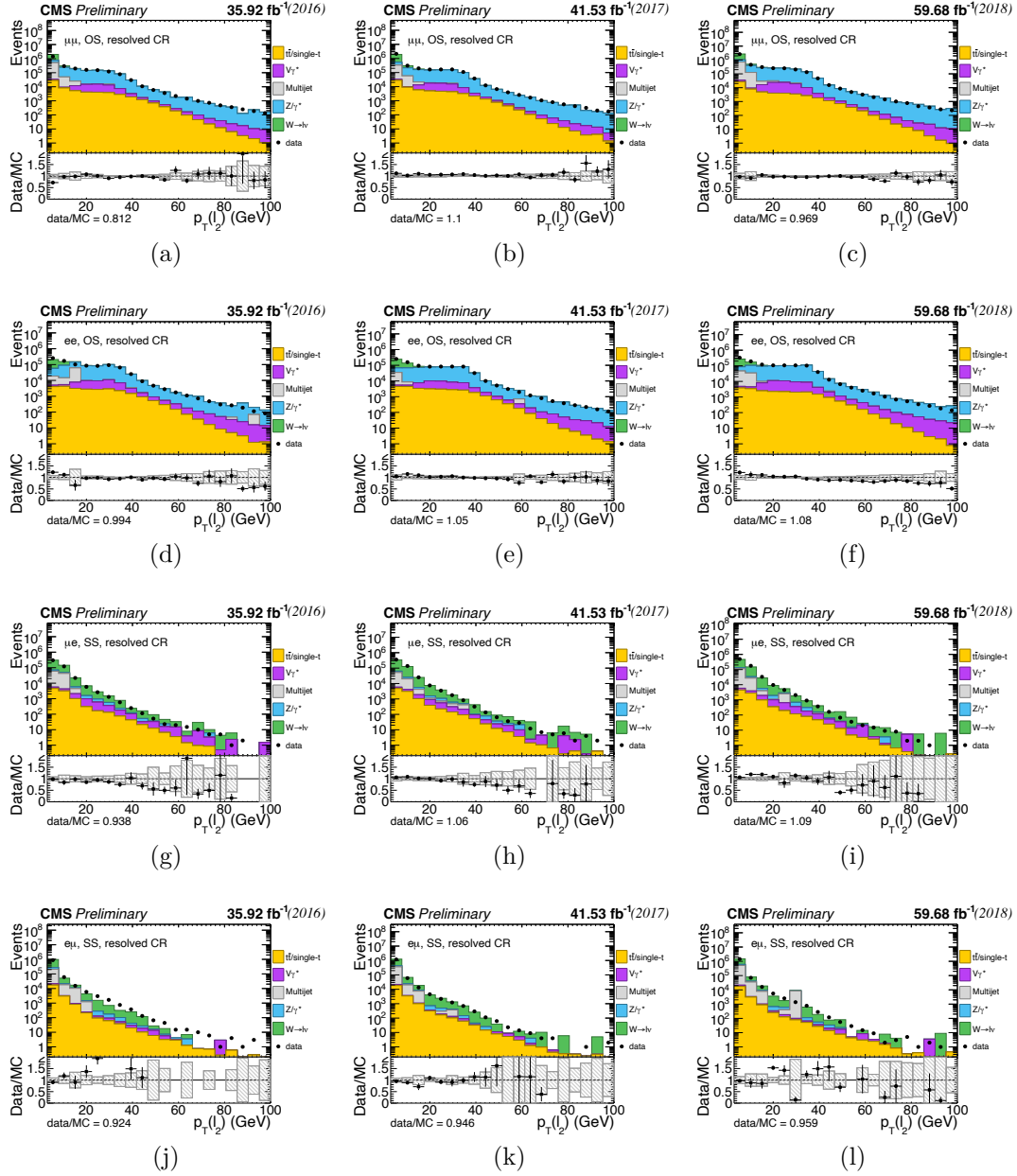


Figure C.7: Distribution of the ℓ_2 transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

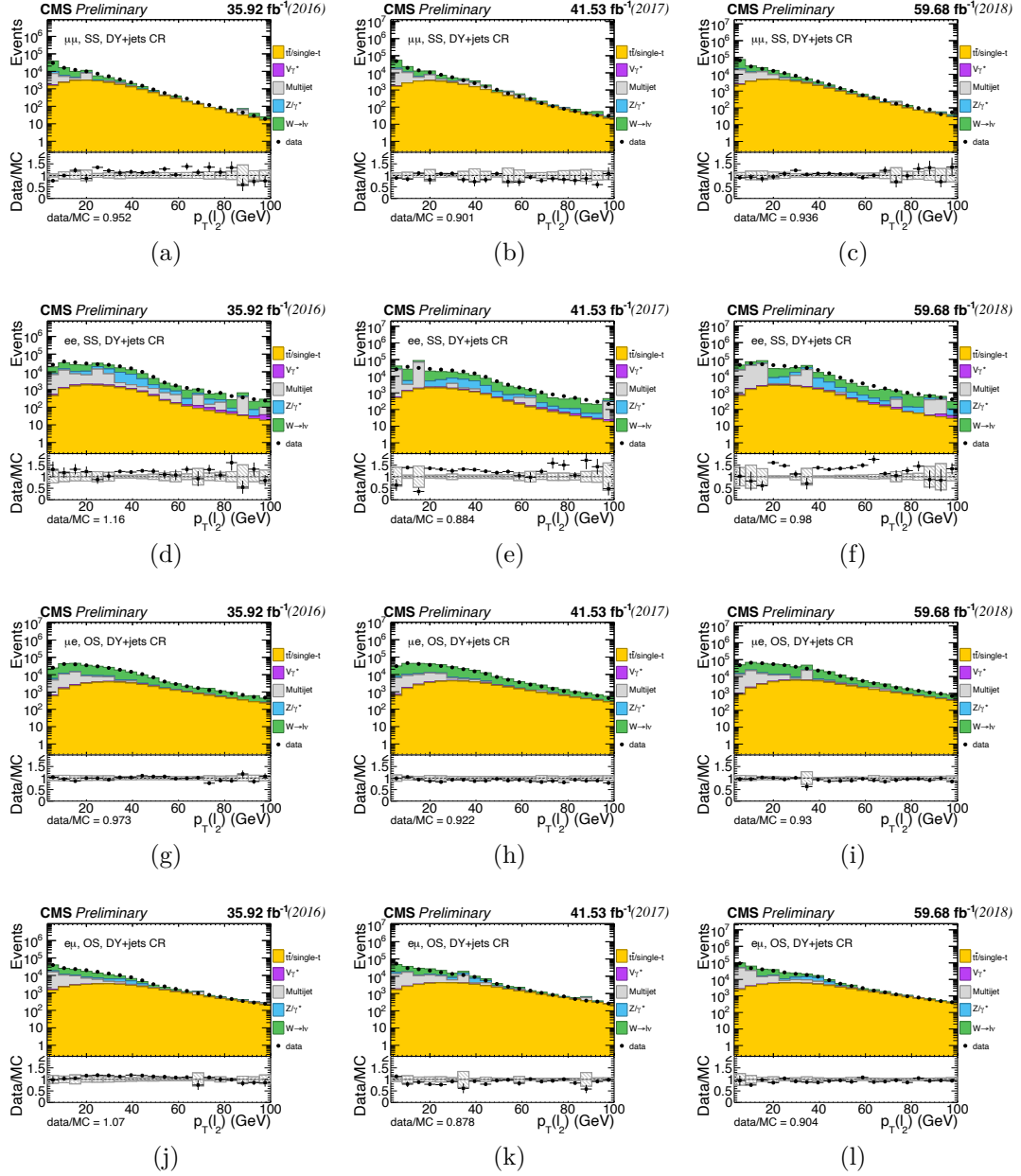


Figure C.8: Distribution of the ℓ_2 momentum in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.

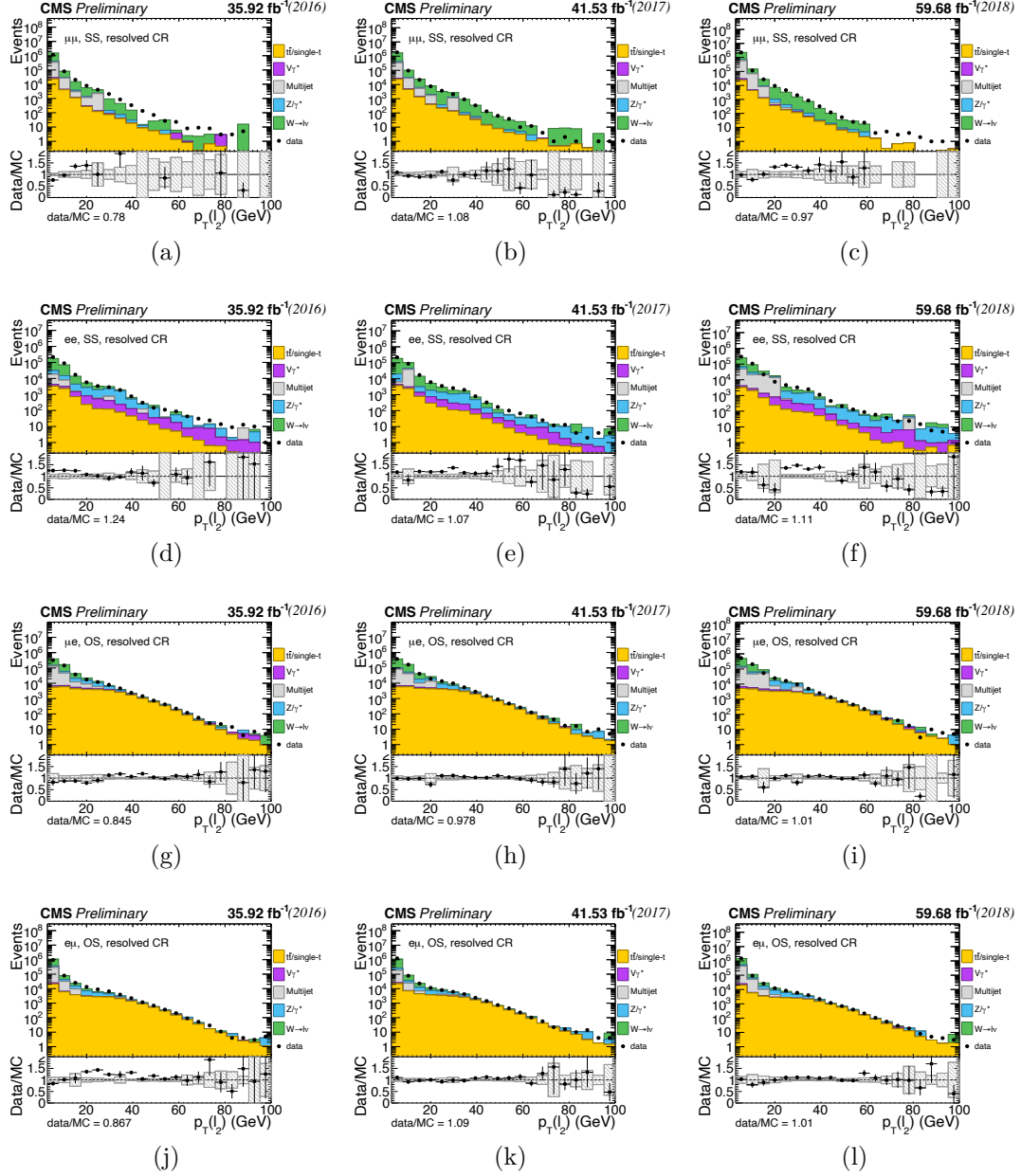


Figure C.9: Distribution of the ℓ_2 transverse momentum in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.

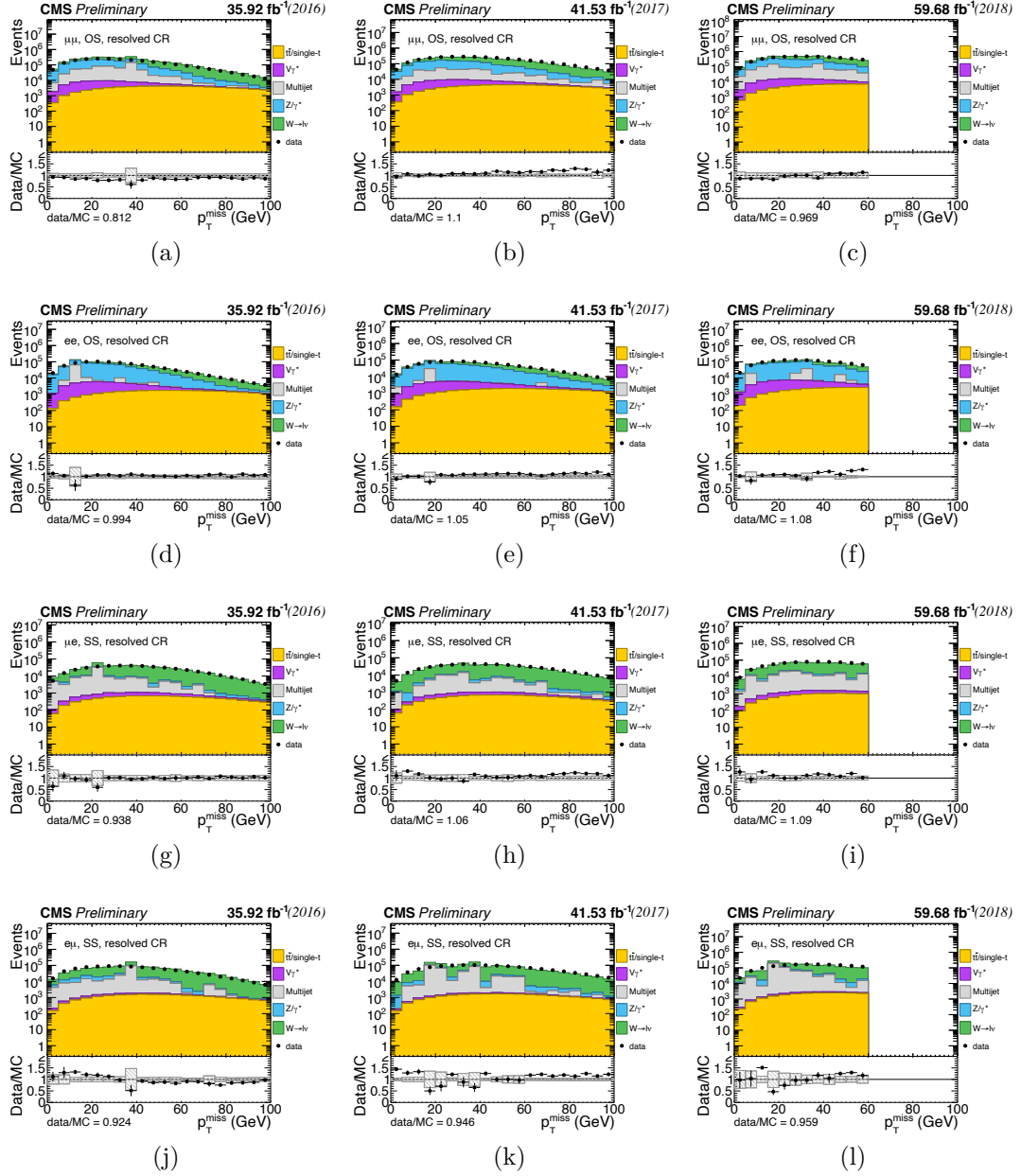


Figure C.10: Distribution of the p_T^{miss} in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

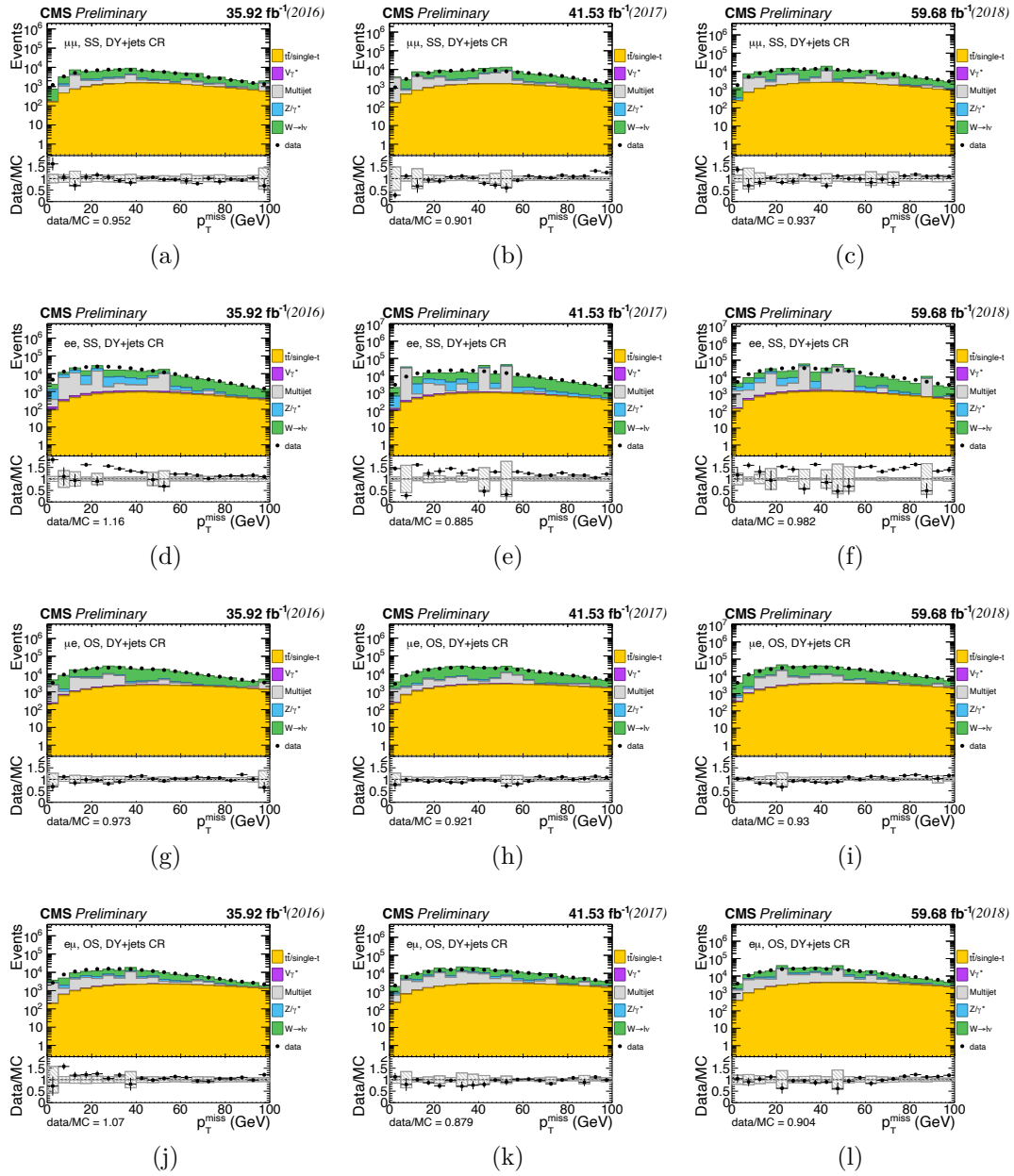


Figure C.11: Distribution of the p_T^{miss} in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.

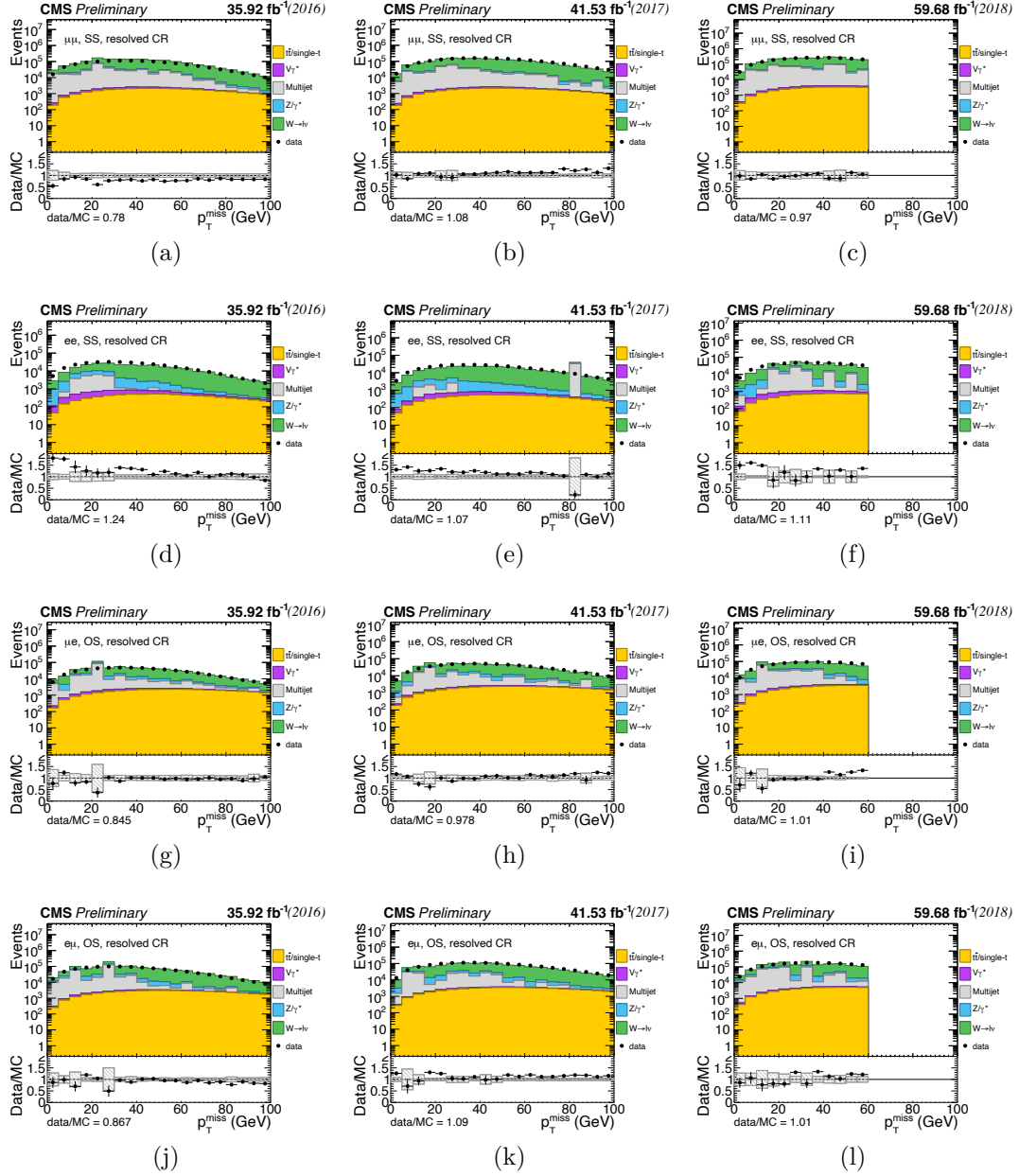


Figure C.12: Distribution of the p_T^{miss} in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the SSSF categories in the first two rows, and the OS mixed flavour categories in the bottom two rows.

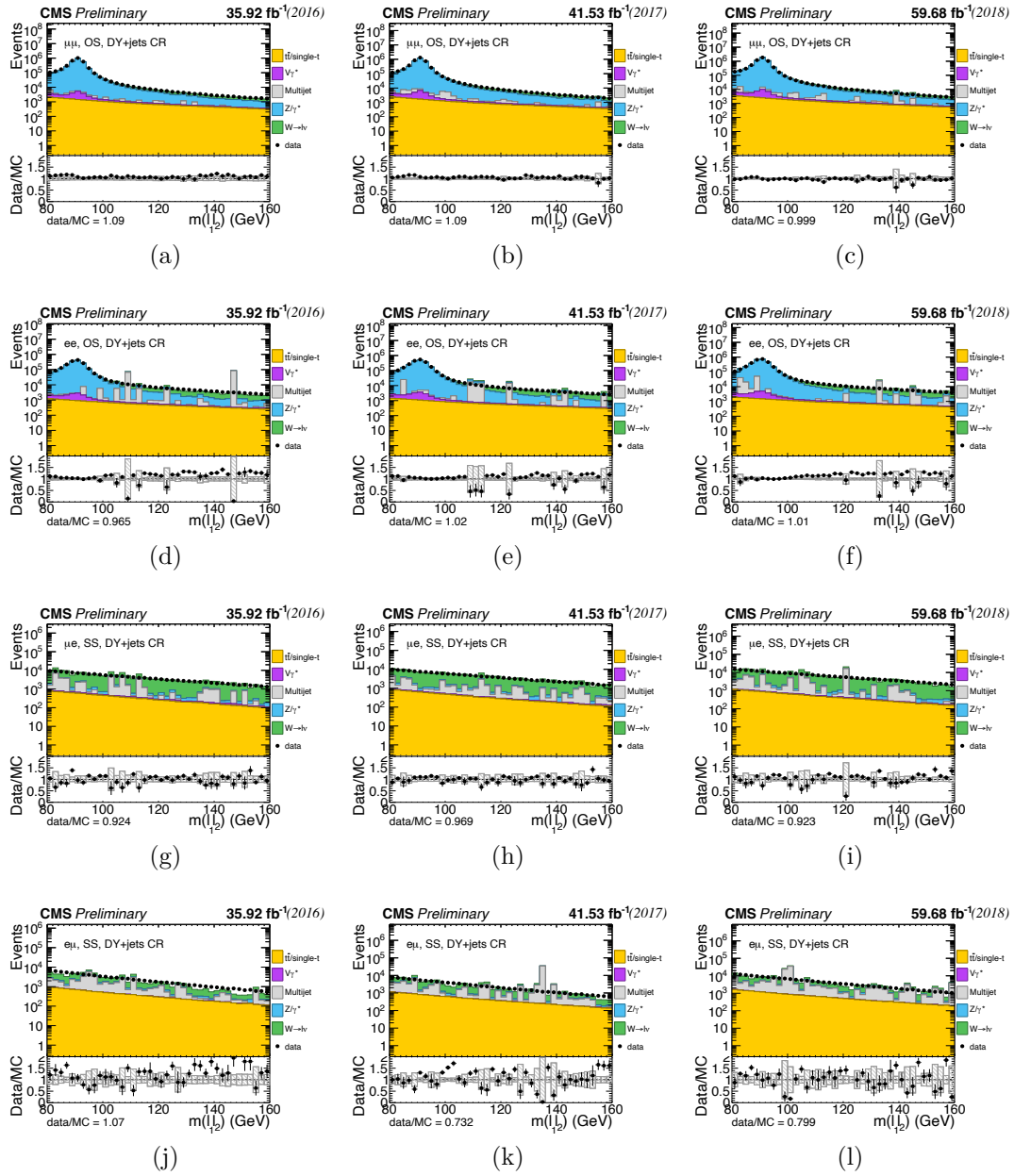


Figure C.13: Distribution of the dilepton mass in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

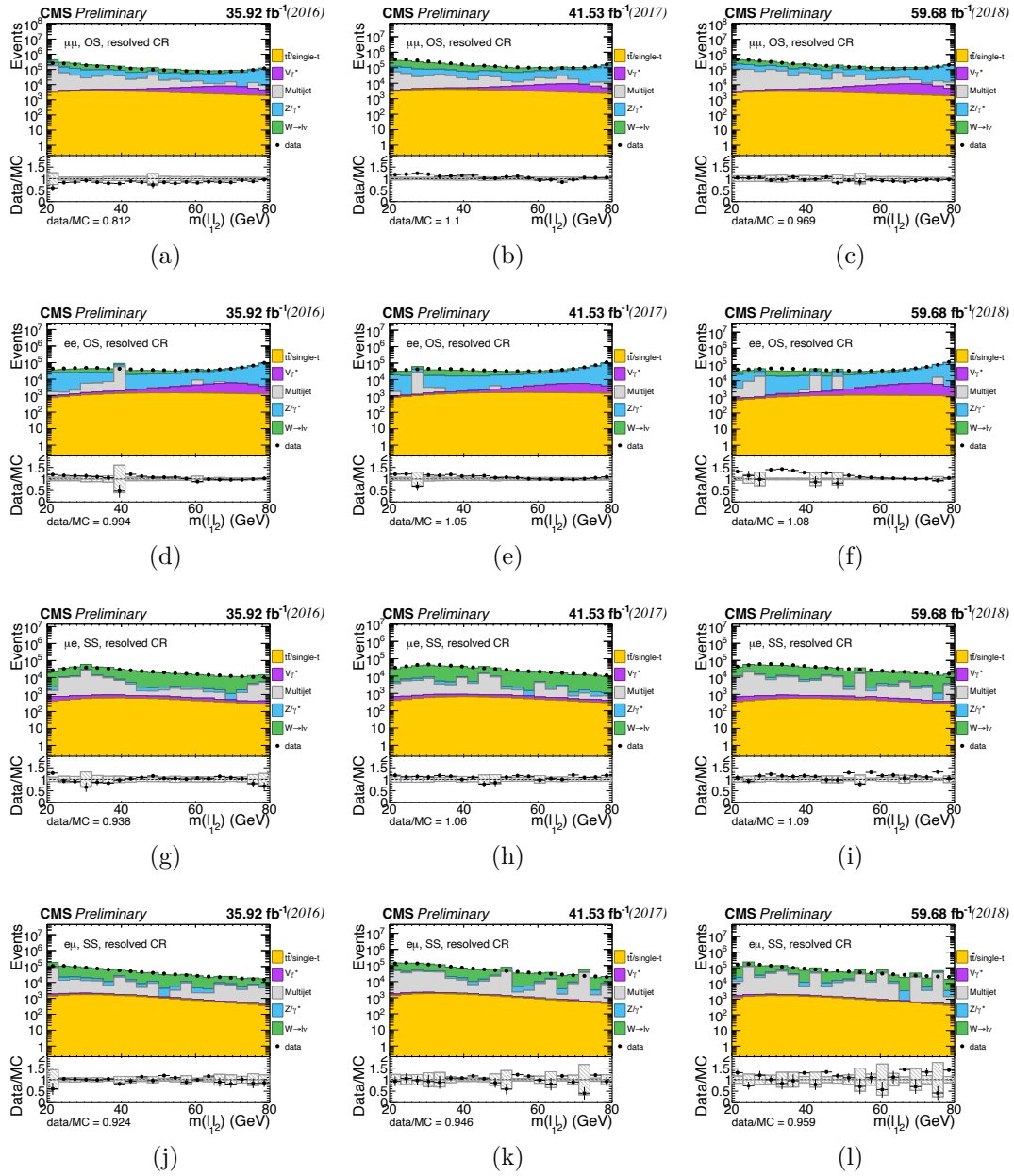


Figure C.14: Distribution of the dilepton mass in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

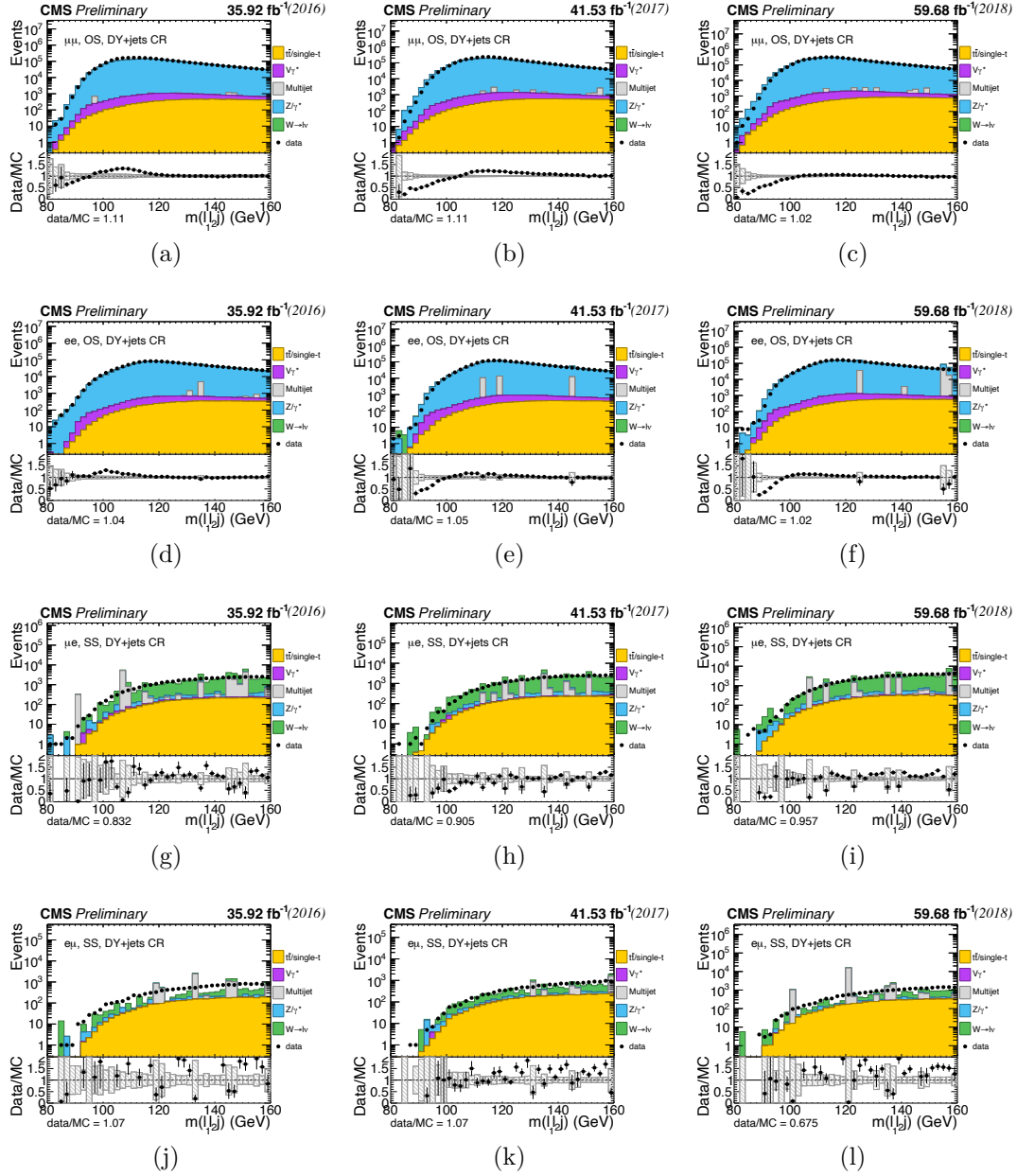


Figure C.15: Distribution of the 3-body invariant mass variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

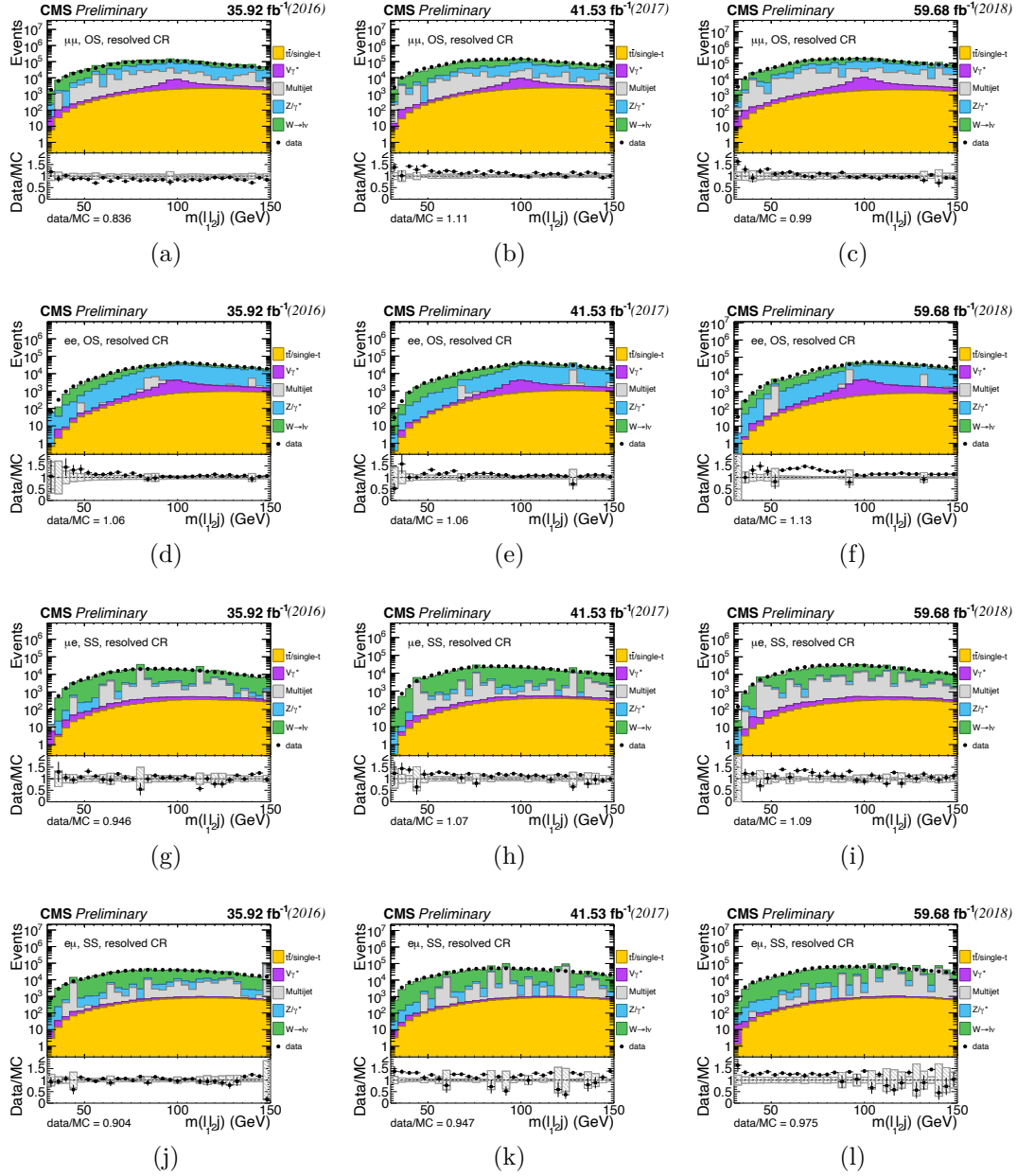


Figure C.16: Distribution of the 3-body invariant mass variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

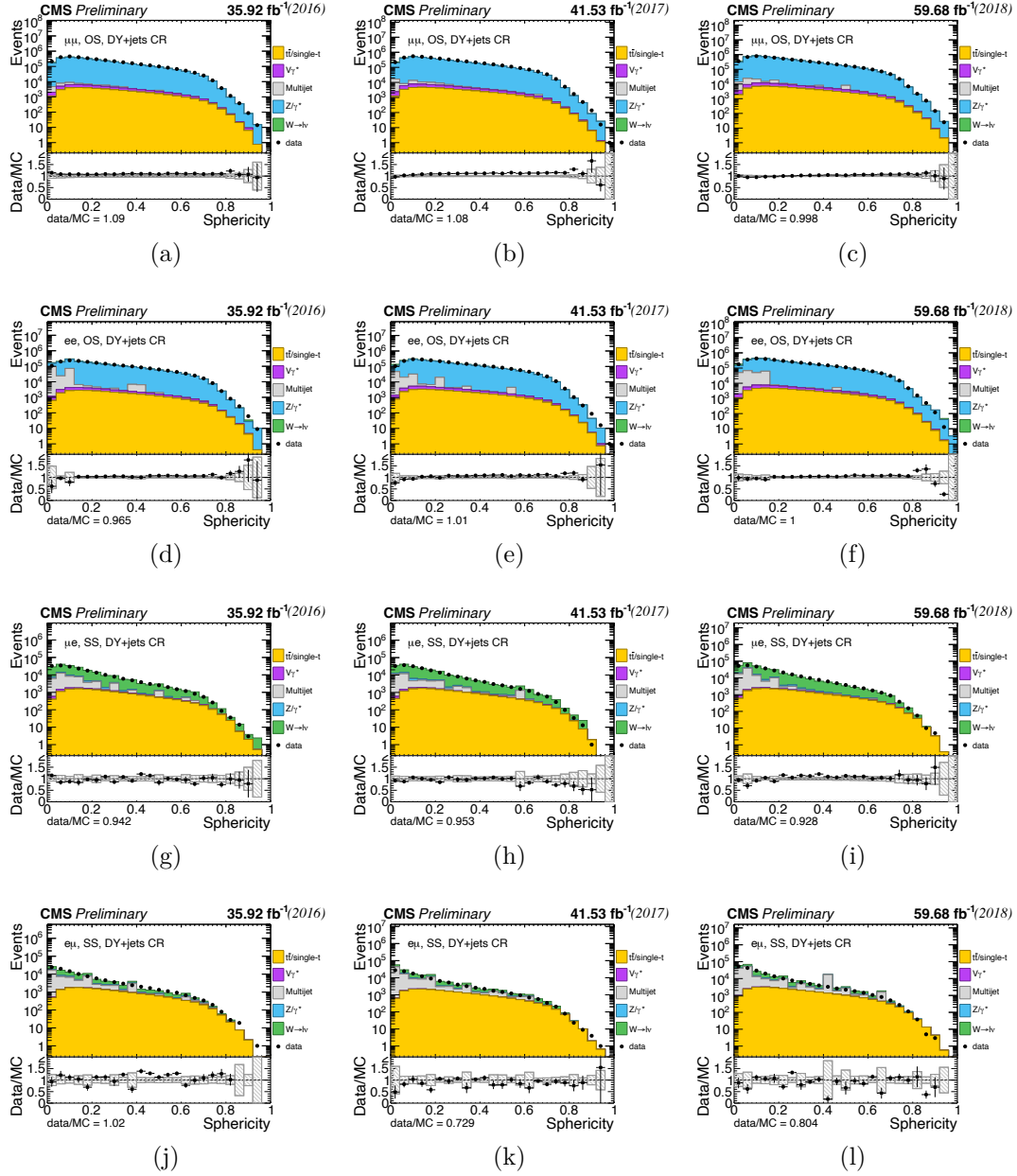


Figure C.17: Distribution of the sphericity event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

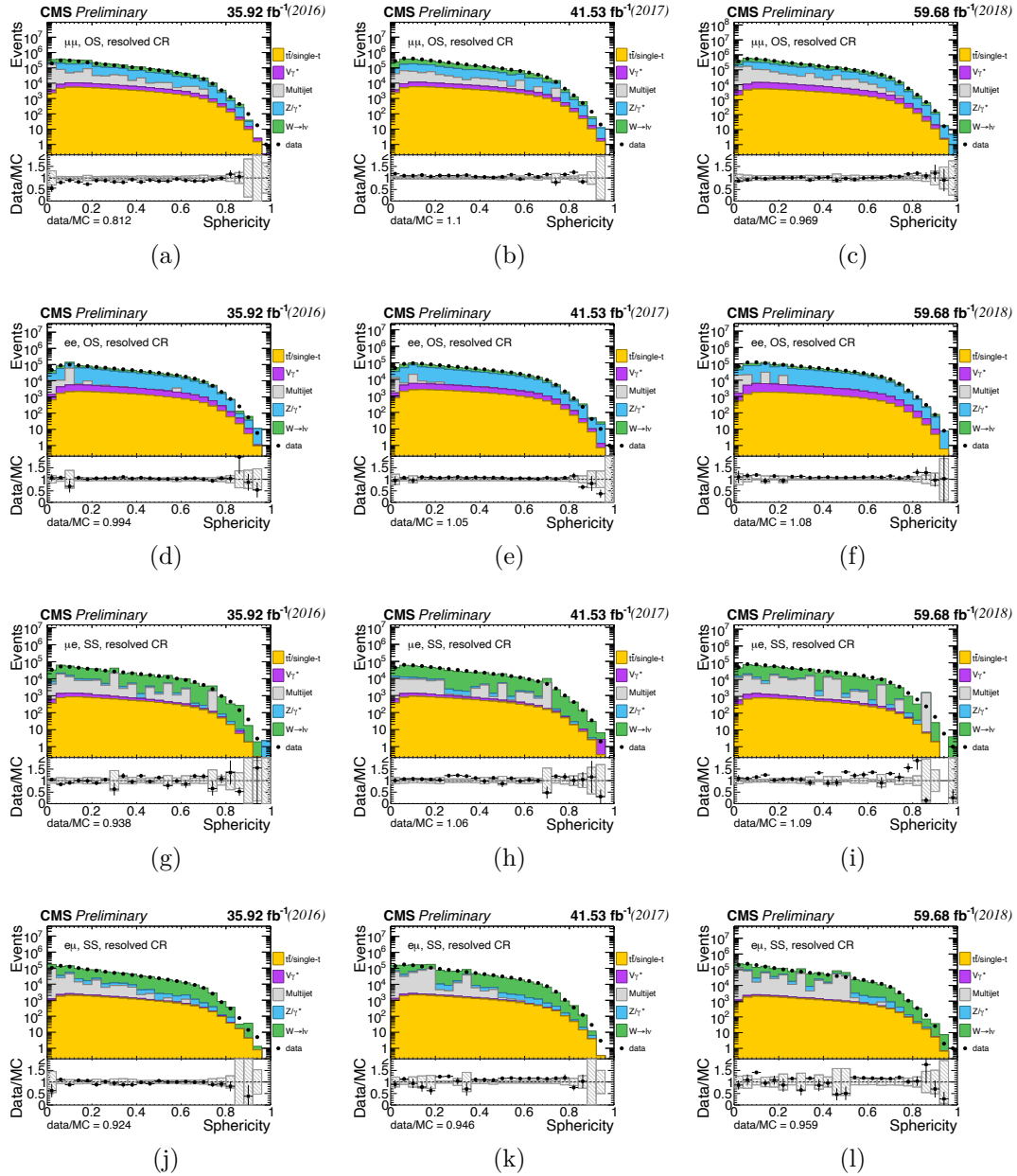


Figure C.18: Distribution of the sphericity event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

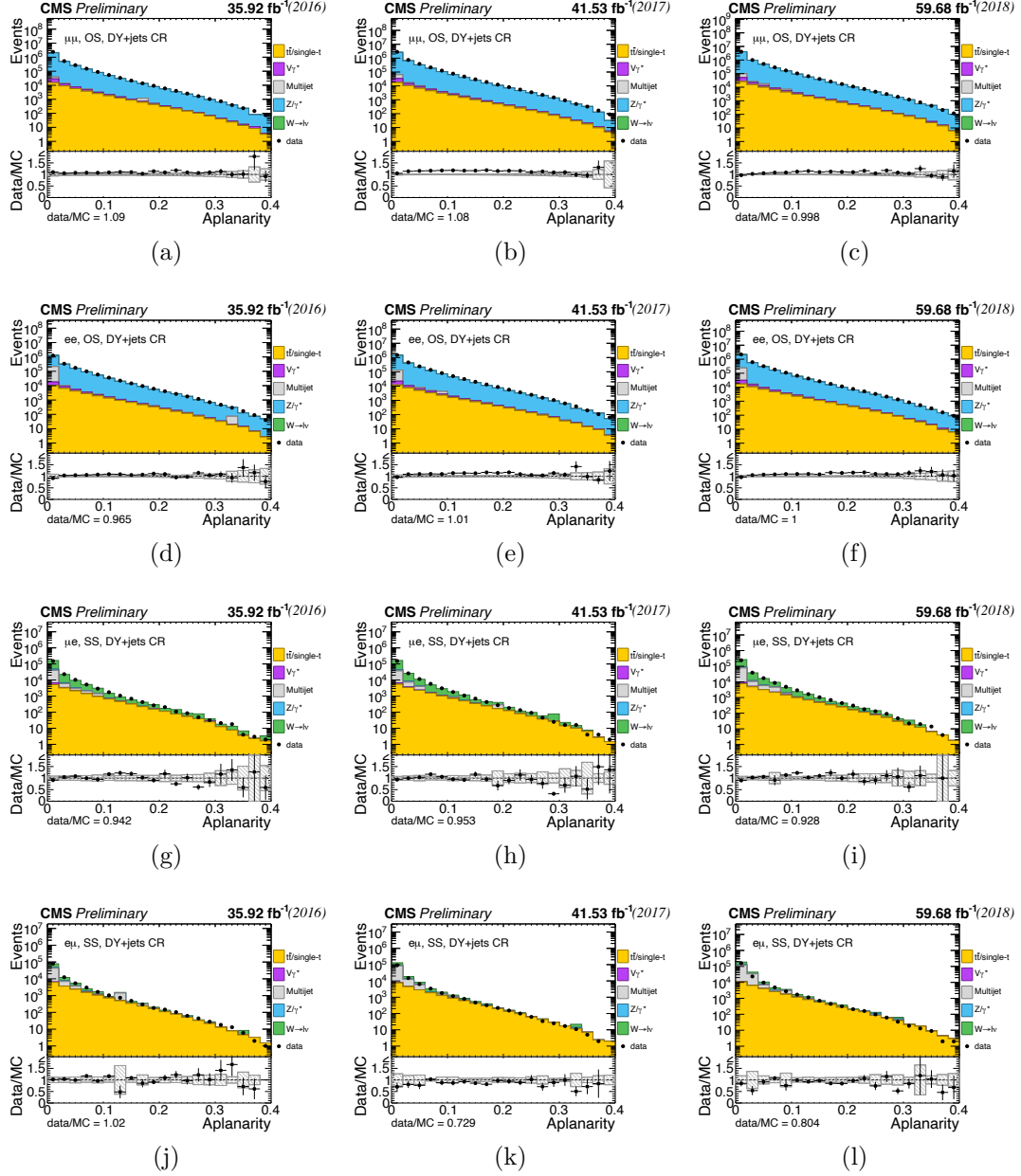


Figure C.19: Distribution of the aplanarity event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

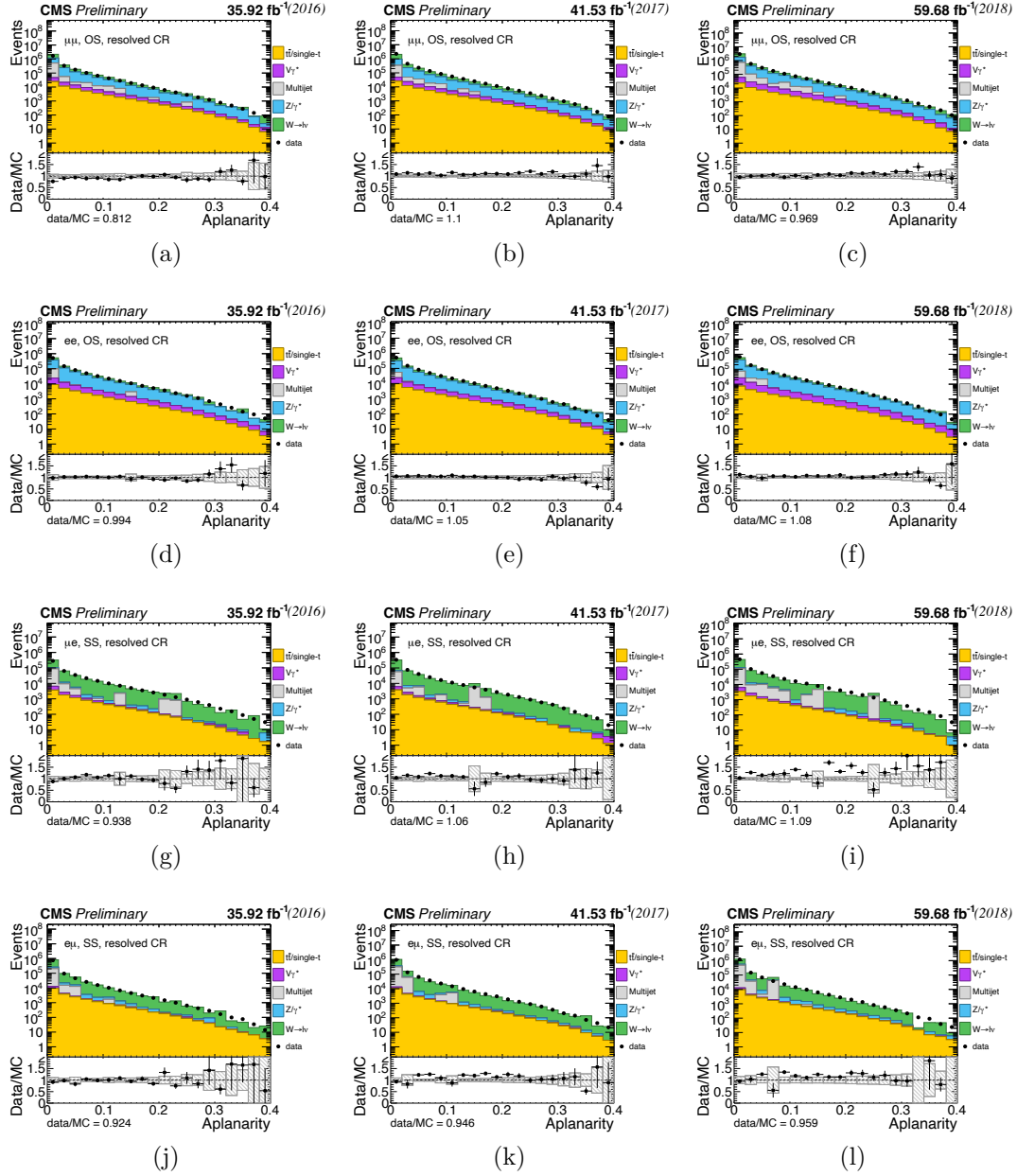


Figure C.20: Distribution of the aplanarity event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

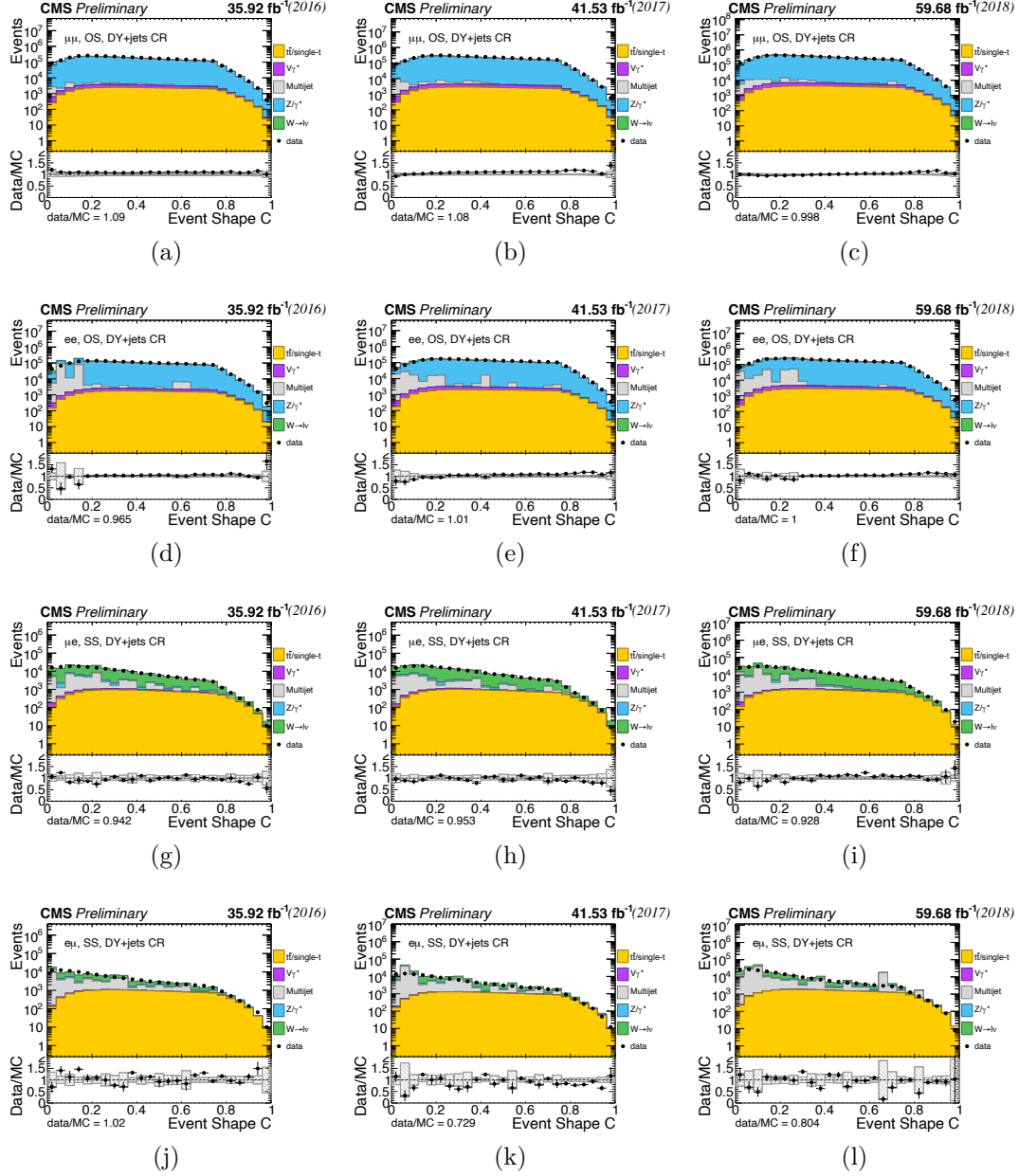


Figure C.21: Distribution of the C-parameter event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

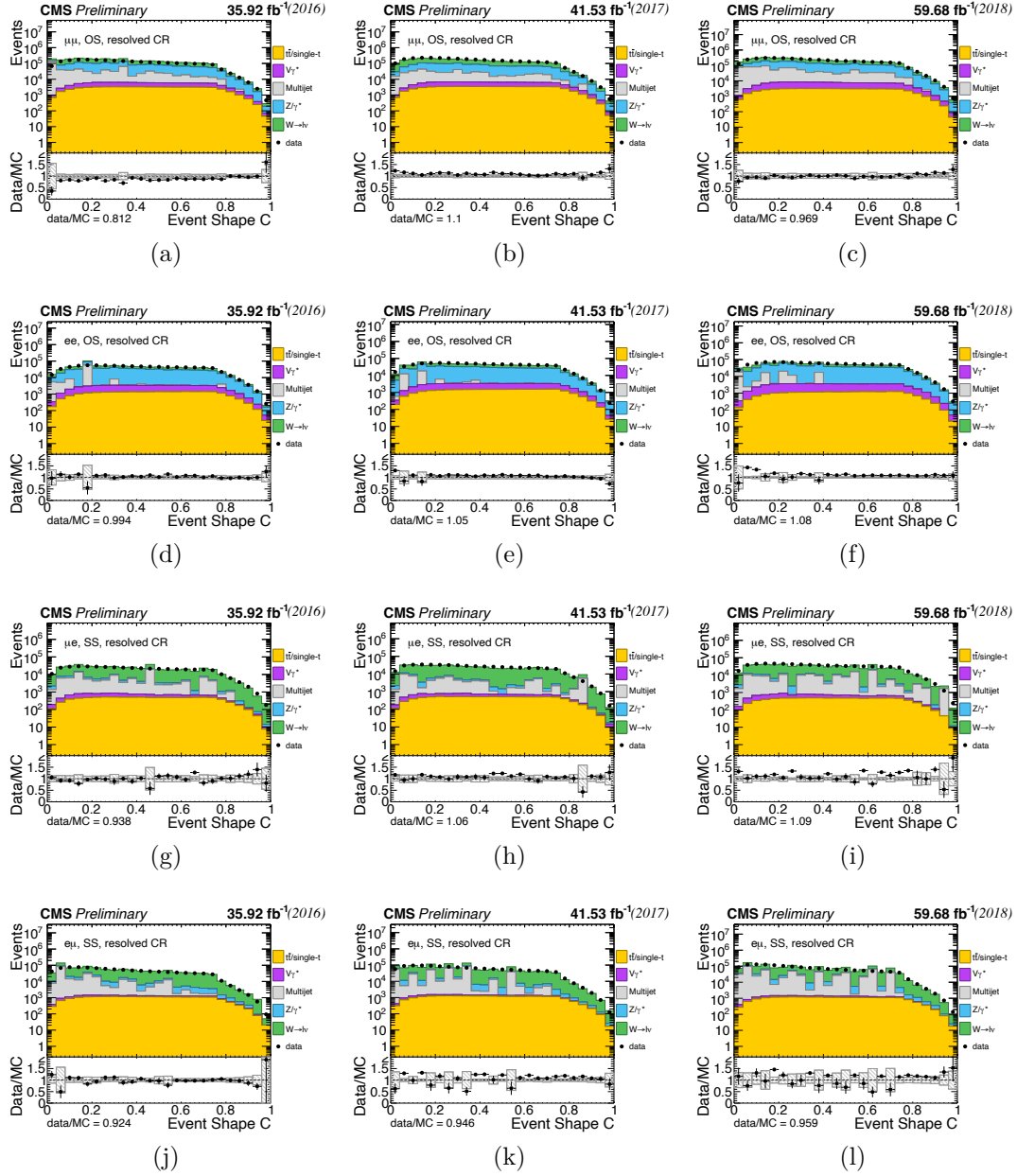


Figure C.22: Distribution of the C-parameter event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

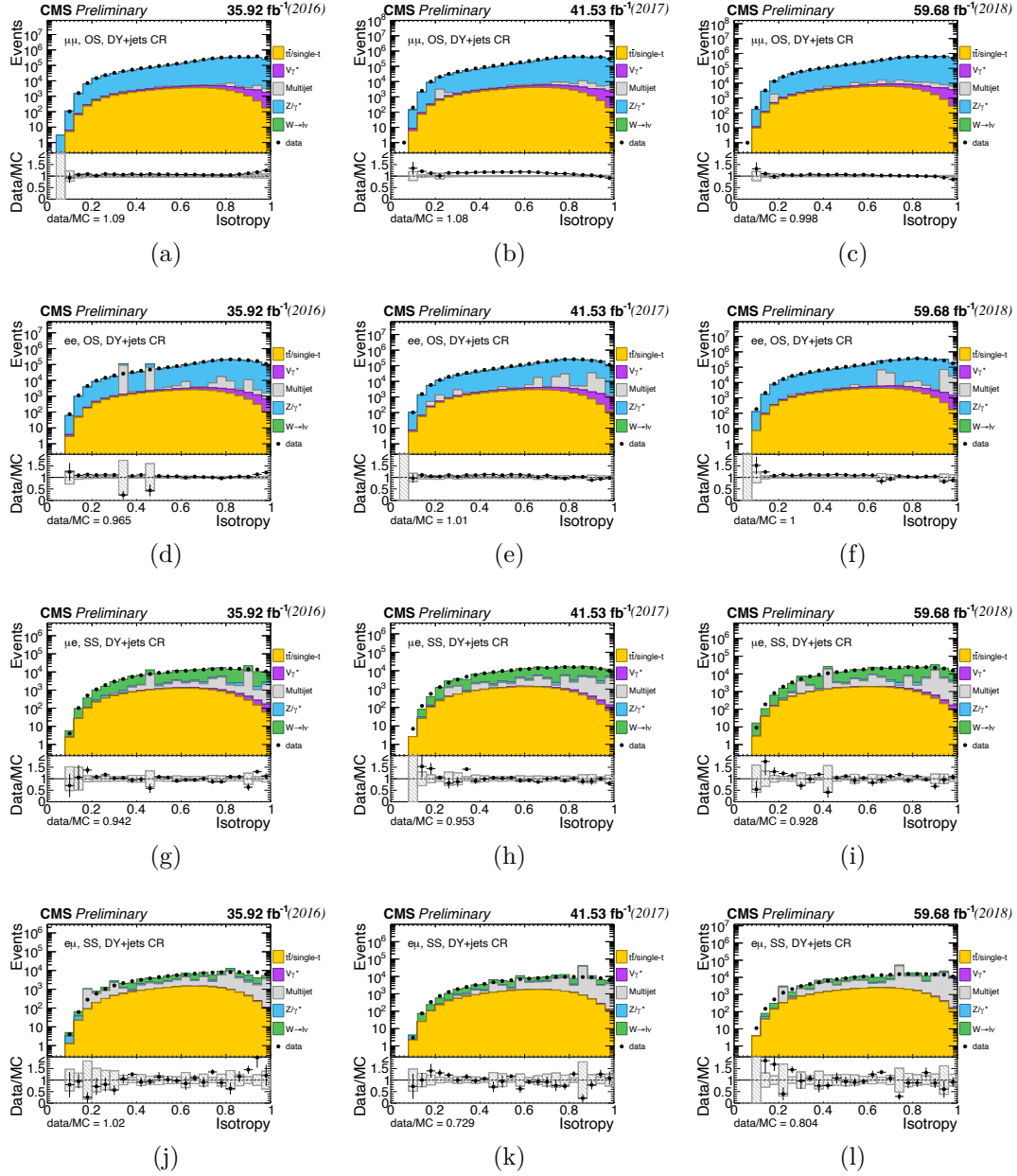


Figure C.23: Distribution of the isotropy event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

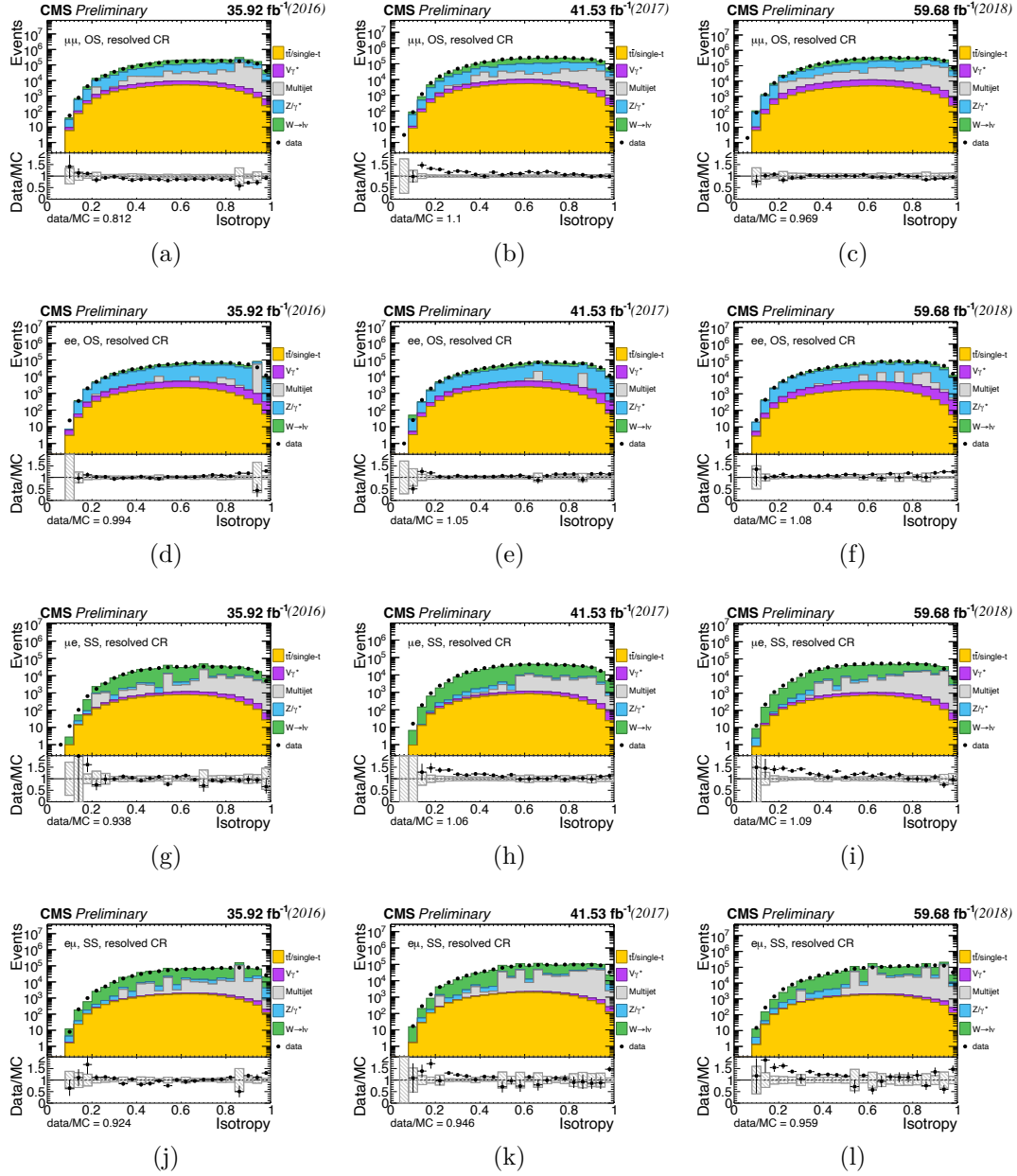


Figure C.24: Distribution of the isotropy event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

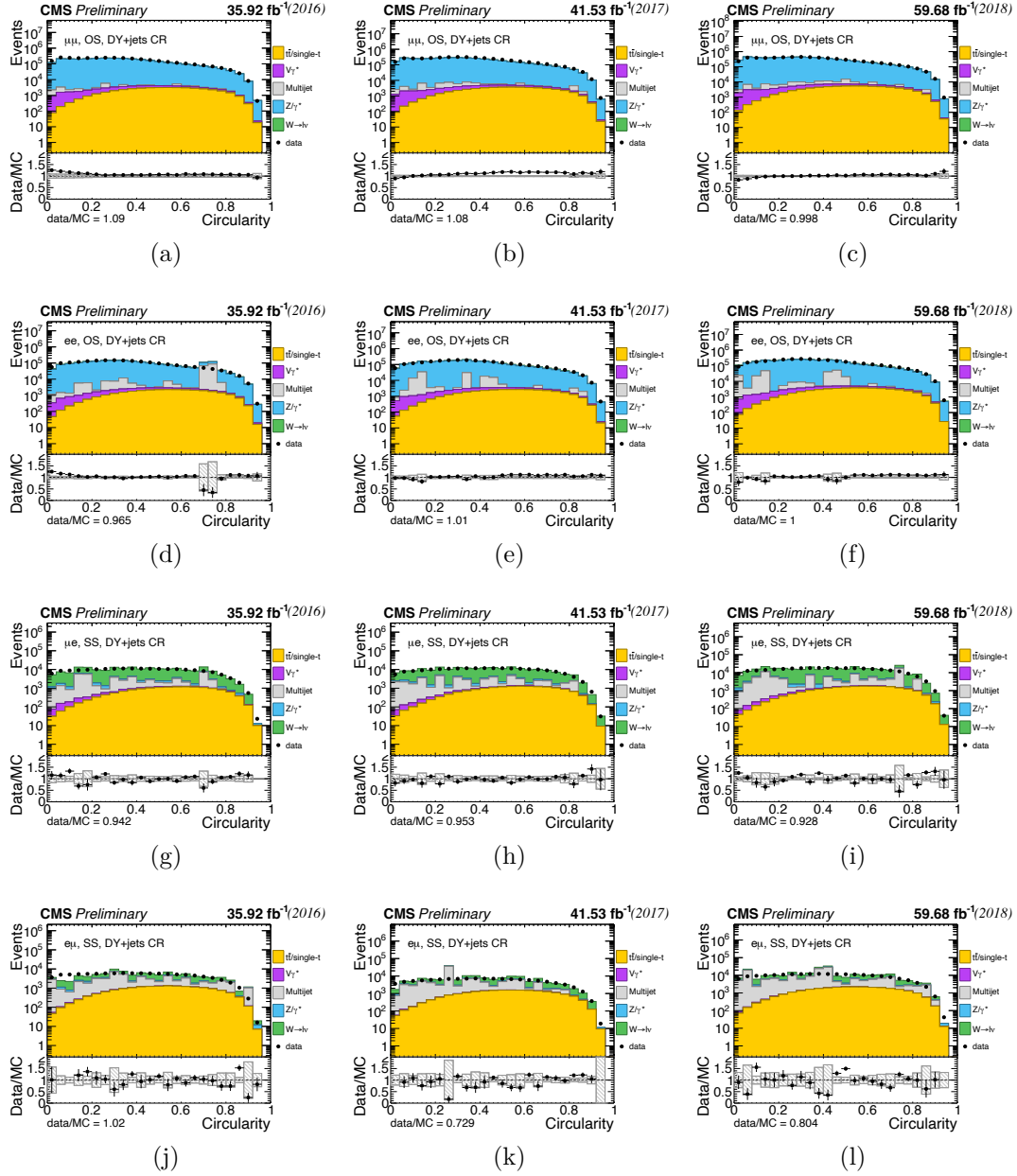


Figure C.25: Distribution of the circularity event shape variable in the DY control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

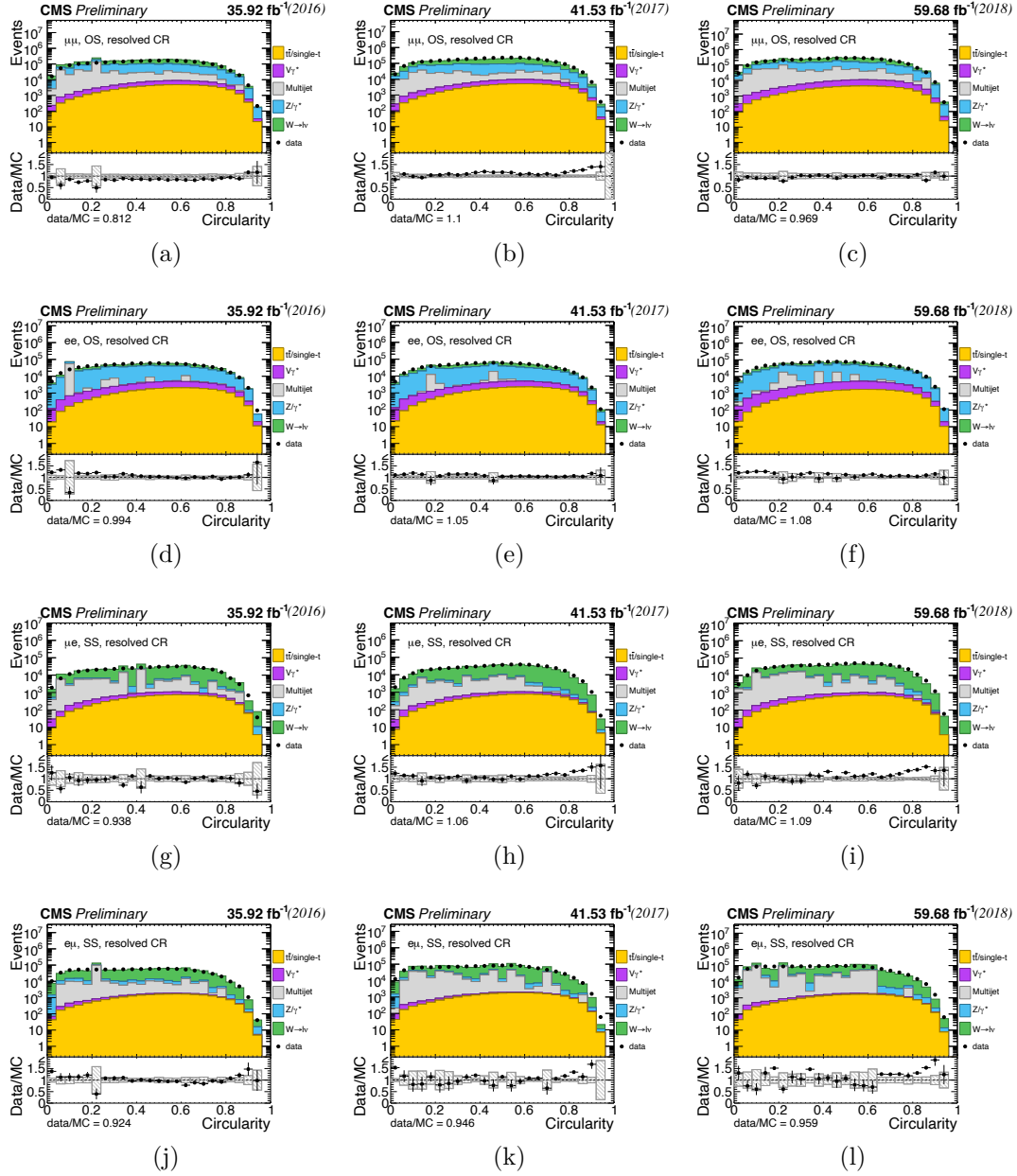


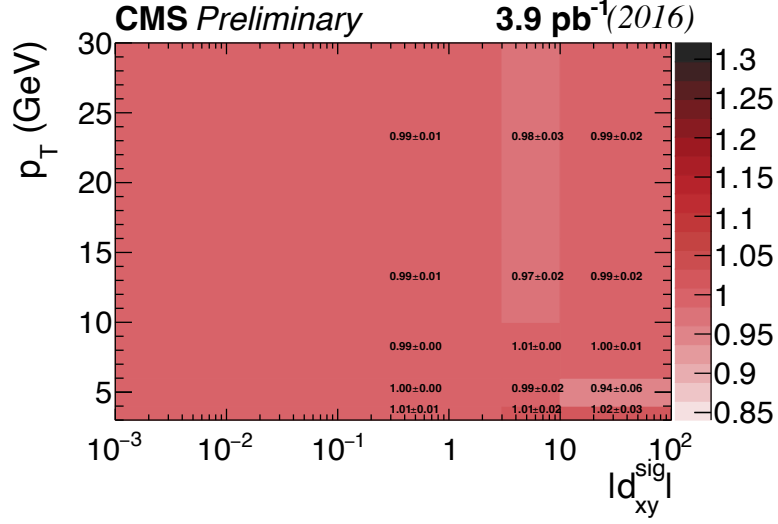
Figure C.26: Distribution of the circularity event shape variable in the resolved control region. The distributions are shown for 2016, 2017 and 2018 scenarios in the left, middle and right columns, respectively. The distributions are shown for the OSSF categories in the first two rows, and the SS mixed flavour categories in the bottom two rows.

Appendix D

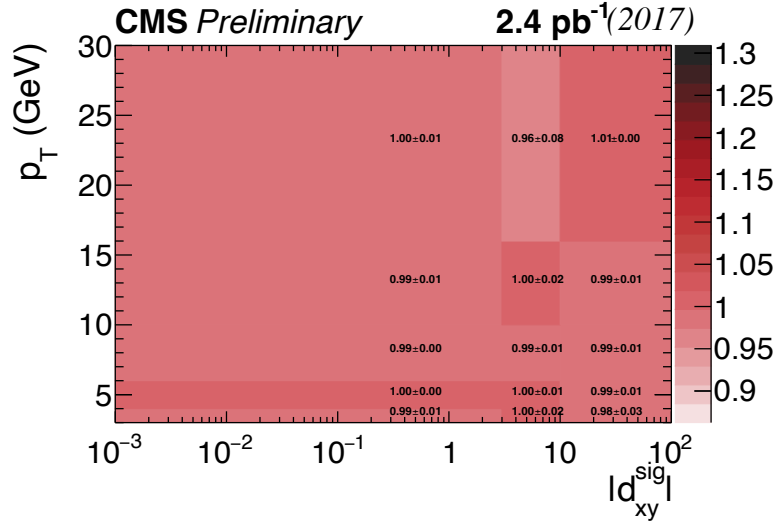
Complementary material to the displaced muon scale factor studies

D.1 Displaced muon ID scale factors

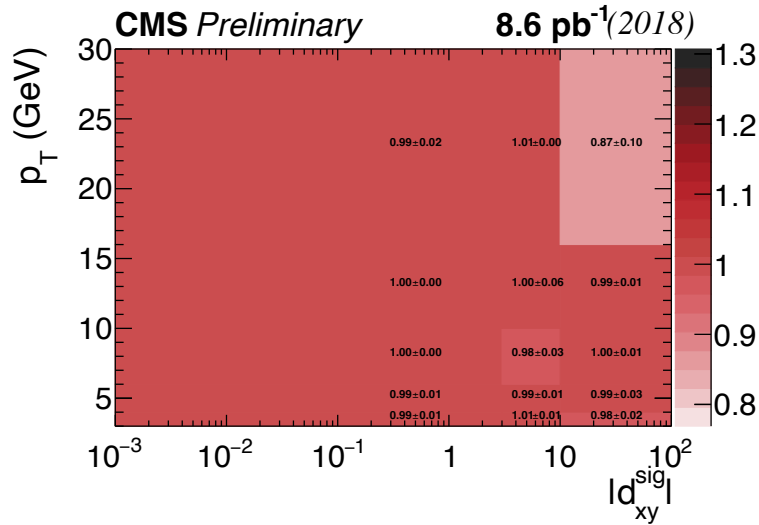
D.2 Displaced muon reconstruction scale factors



(a)

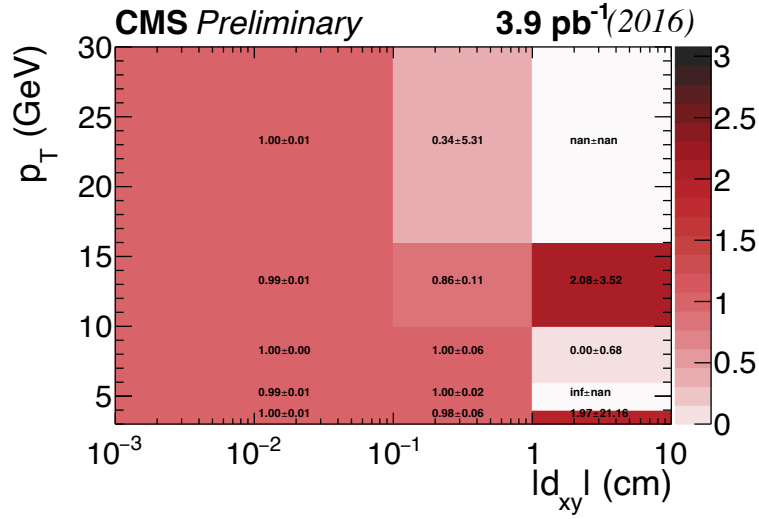


(b)

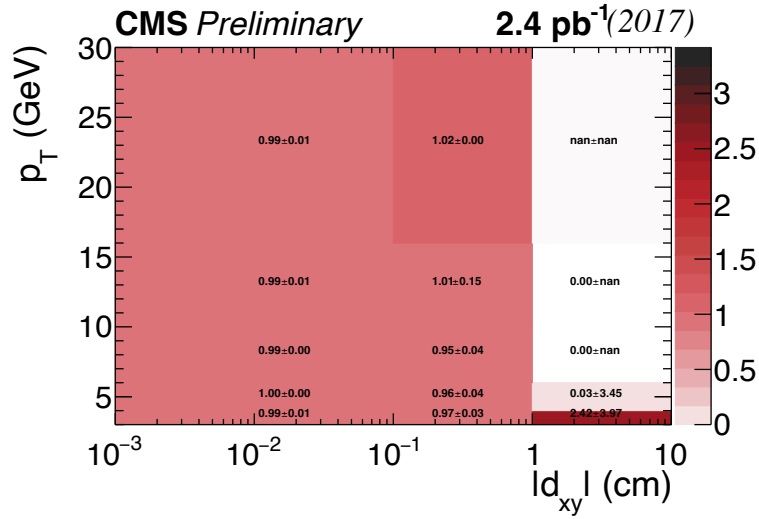


(c)

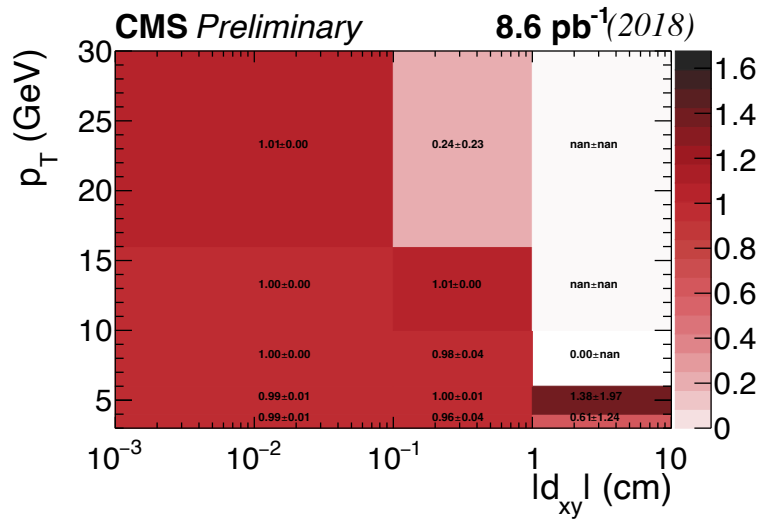
Figure D.1: Scale factor values evaluated for displaced muon loose identification criteria as a function of the muon p_T and $|d_{xy}^{sig}|$ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events. These values have been implemented in the analysis framework as correction scale factors.



(a)

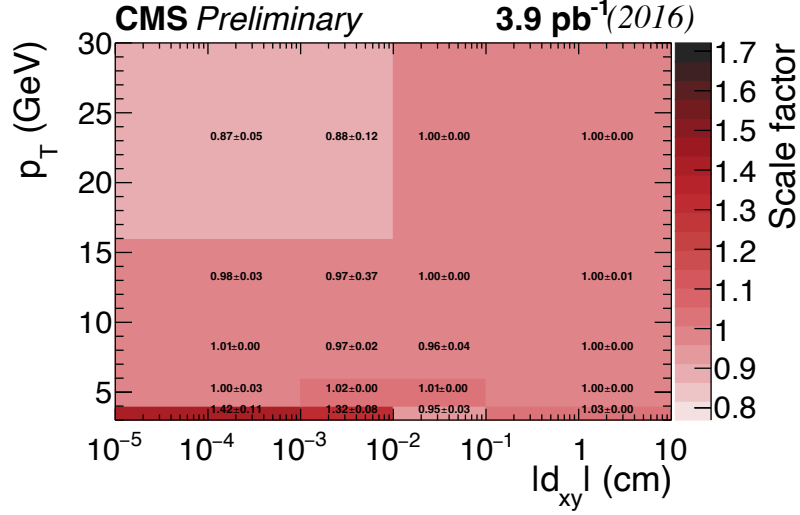


(b)

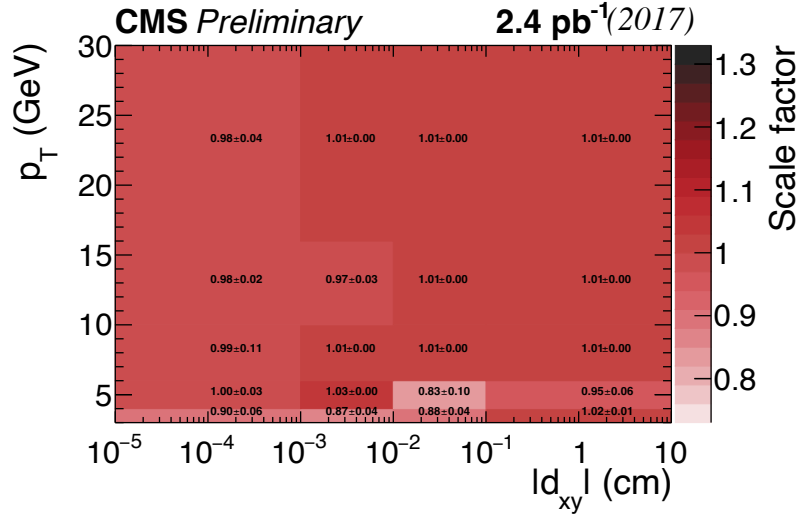


(c)

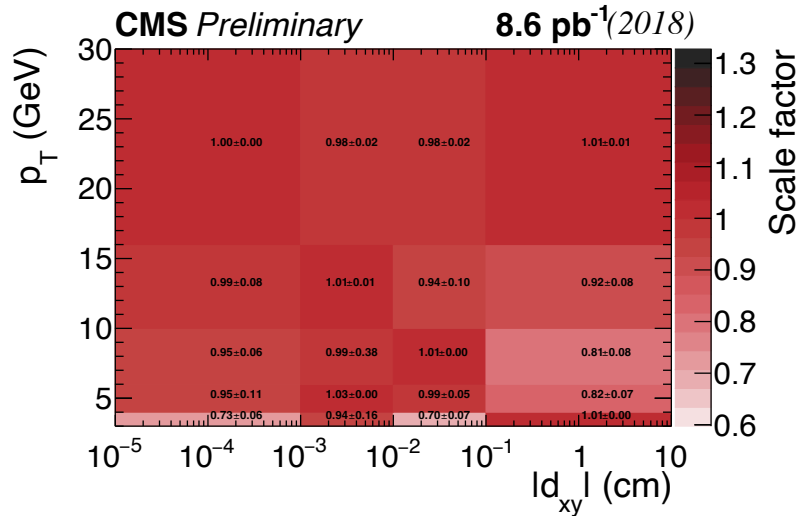
Figure D.2: Scale factor values evaluated for displaced muon loose identification criteria as a function of the muon p_T and $|d_{xy}|$ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events.



(a)



(b)



(c)

Figure D.3: Scale factor values evaluated for displaced muon reconstruction as a function of the muon p_T and $|d_{xy}|$ obtained via the tag-and-probe method in $B^\pm \rightarrow J/\psi K^\pm \rightarrow \mu^- \mu^+ K^\pm$ events. These scale factors are eventually propagated in the analysis framework as corrections.

Appendix E

Complementary material to the displaced electron ID scale factor studies

E.1 Data to MC agreement plots in the 3-lepton CR

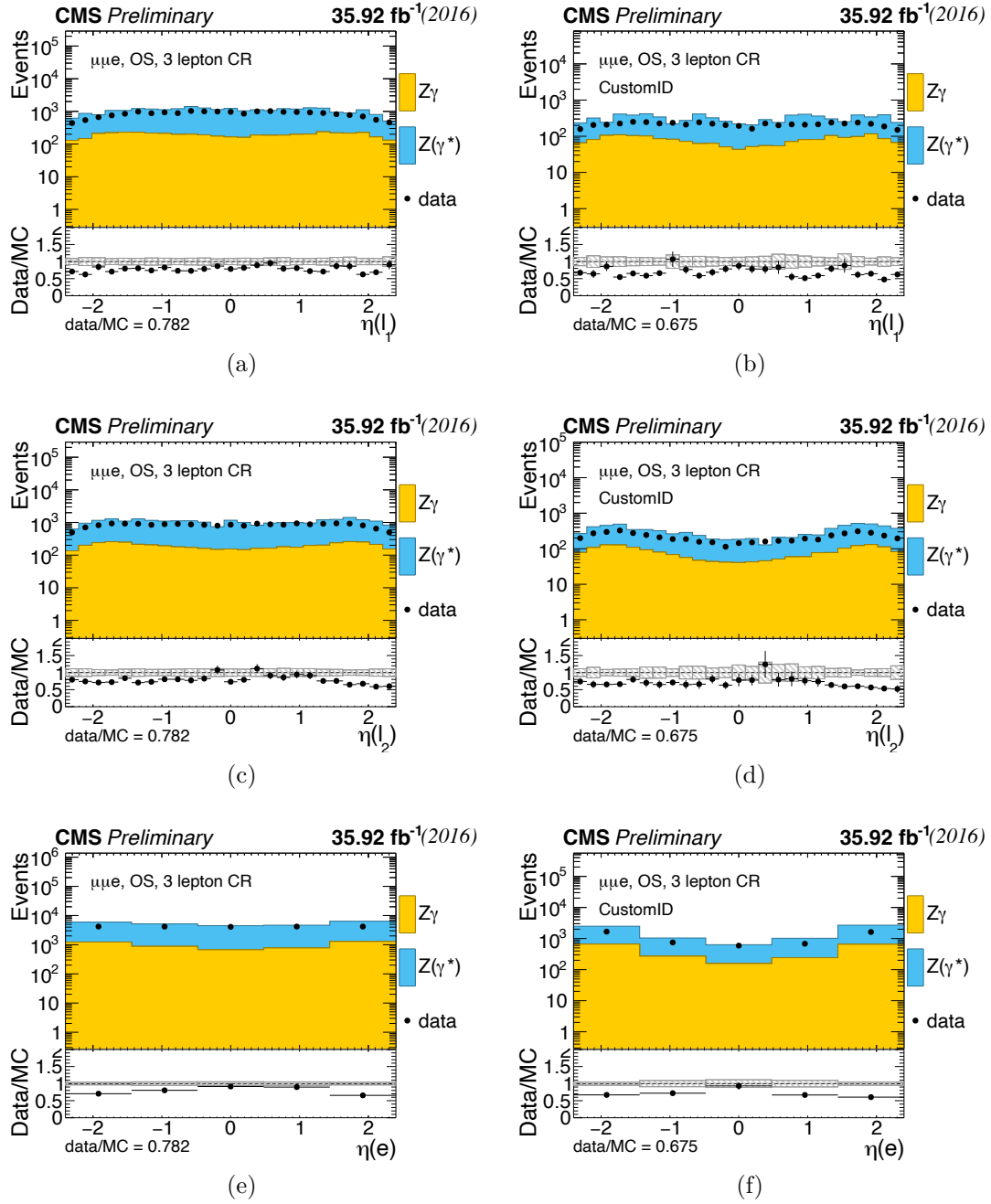


Figure E.1: Data to MC agreement plots in the 3-lepton CR for the leading, subleading and trailing lepton η distributions. Plots on the left do not have custom electron ID applied whereas those on the right do.

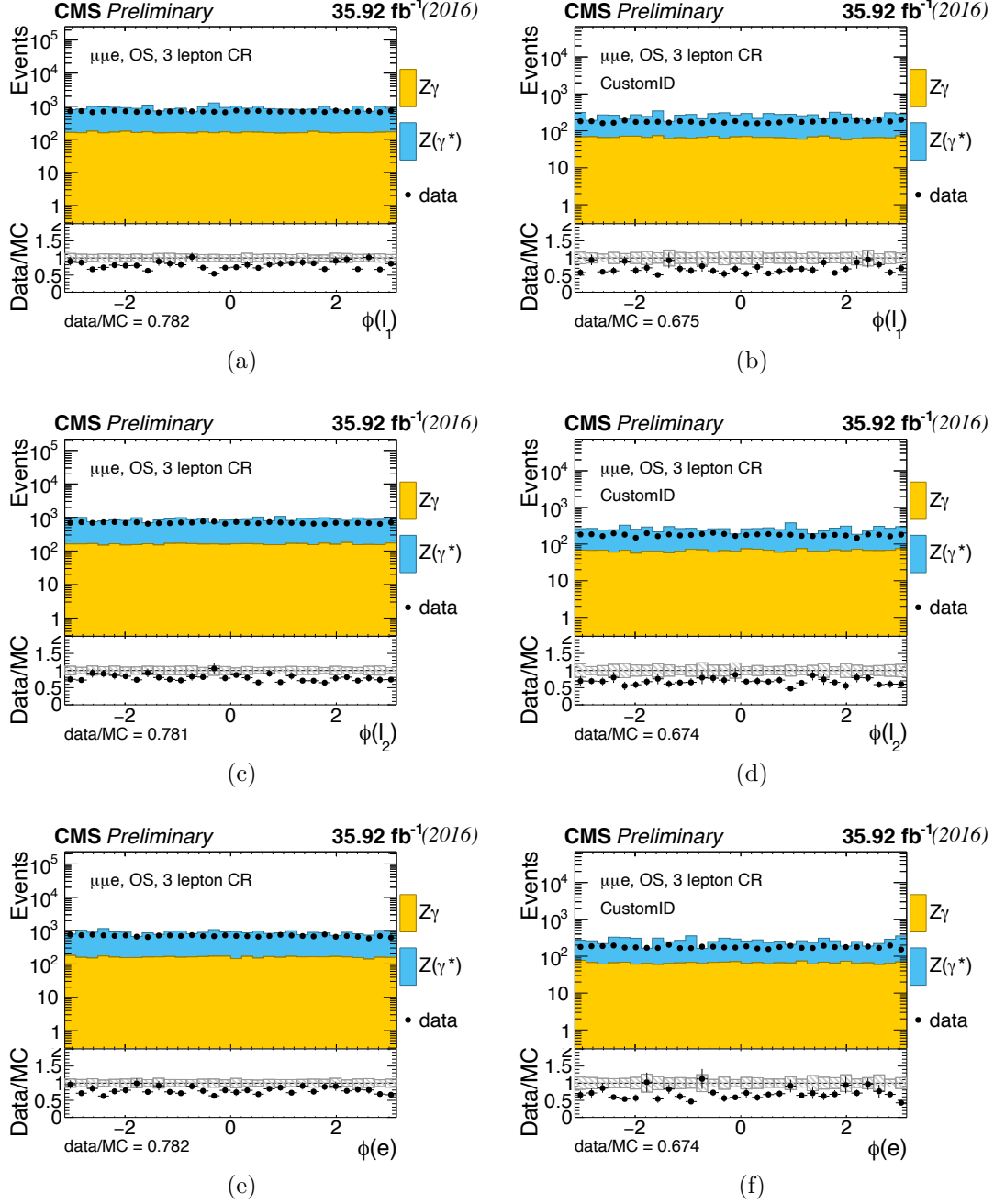


Figure E.2: Data to MC agreement plots in the 3-lepton CR for the leading, subleading and trailing lepton ϕ distributions. Plots on the left do not have custom electron ID applied whereas those on the right do.

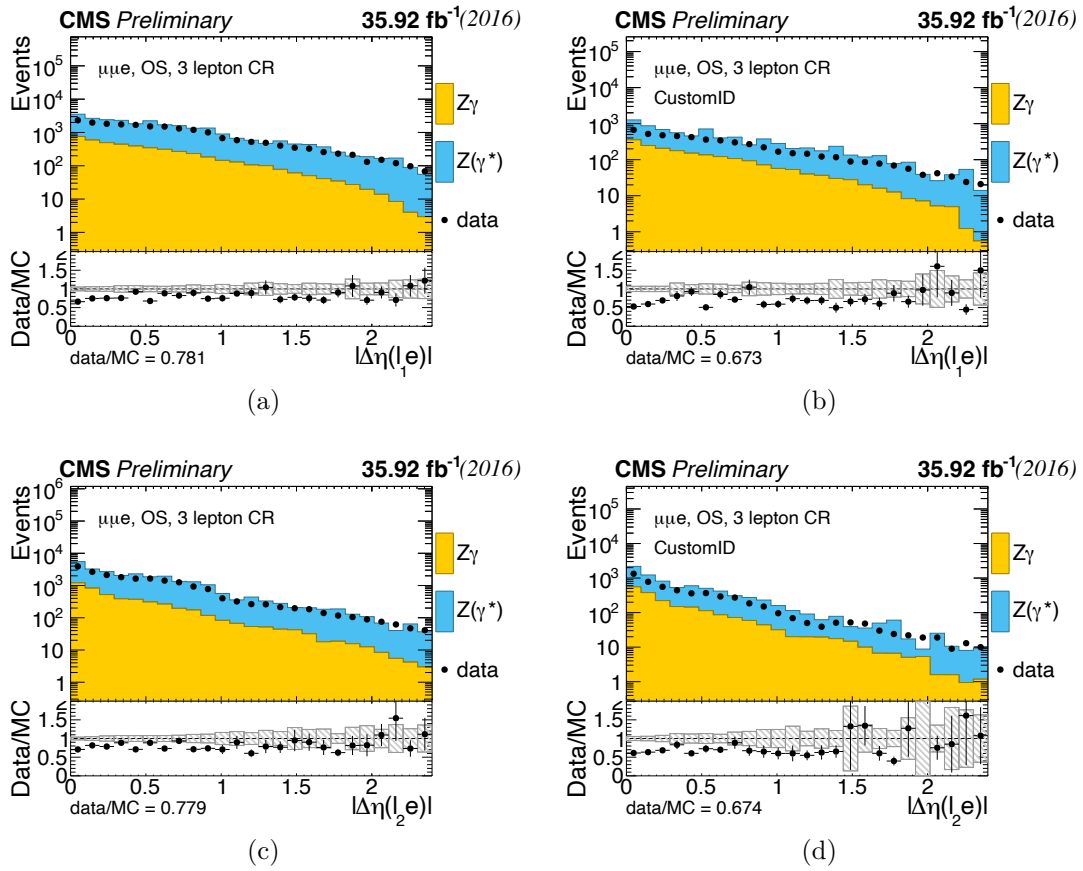


Figure E.3: Data to MC agreement plots in the 3-lepton CR for $\Delta\eta(\ell_1 e)$ and $\Delta\eta(\ell_2 e)$. Plots on the left do not have custom electron ID applied whereas those on the right do.

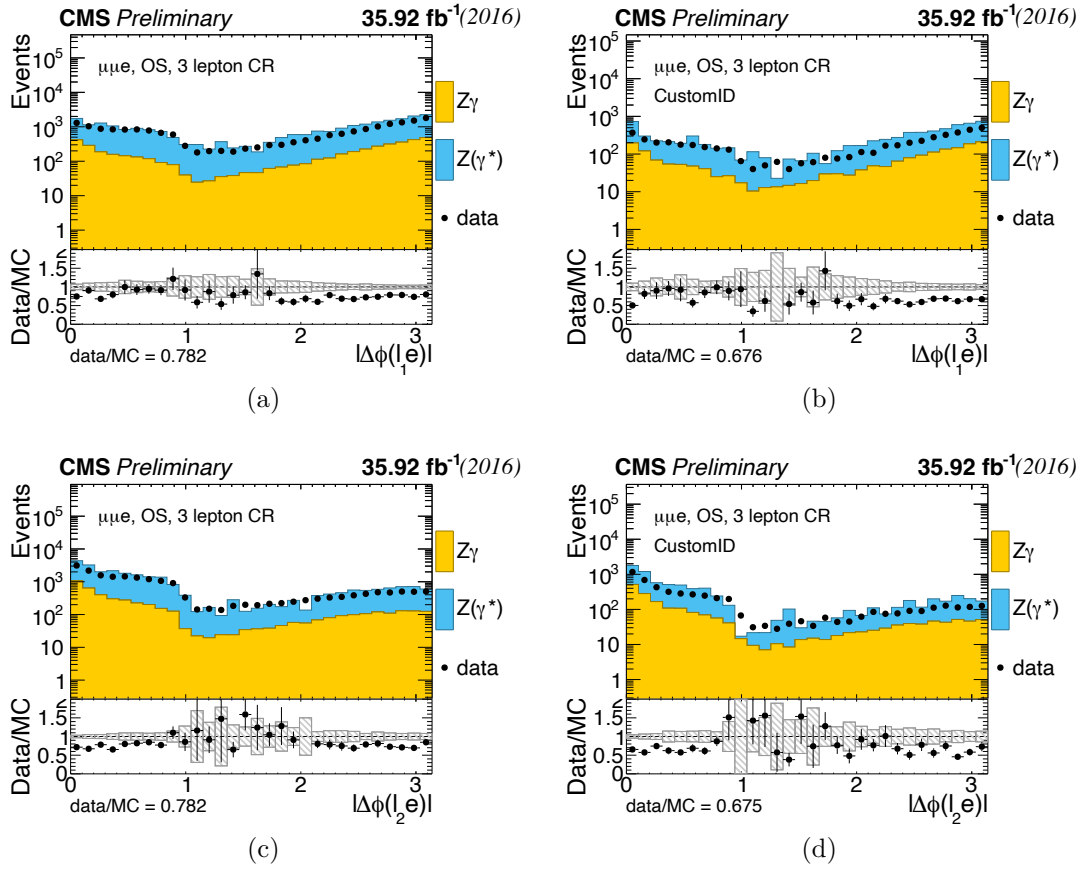


Figure E.4: Data to MC agreement plots in the 3-lepton CR for $\Delta\phi(\ell_1 e)$ and $\Delta\phi(\ell_2 e)$. Plots on the left do not have custom electron ID applied whereas those on the right do.

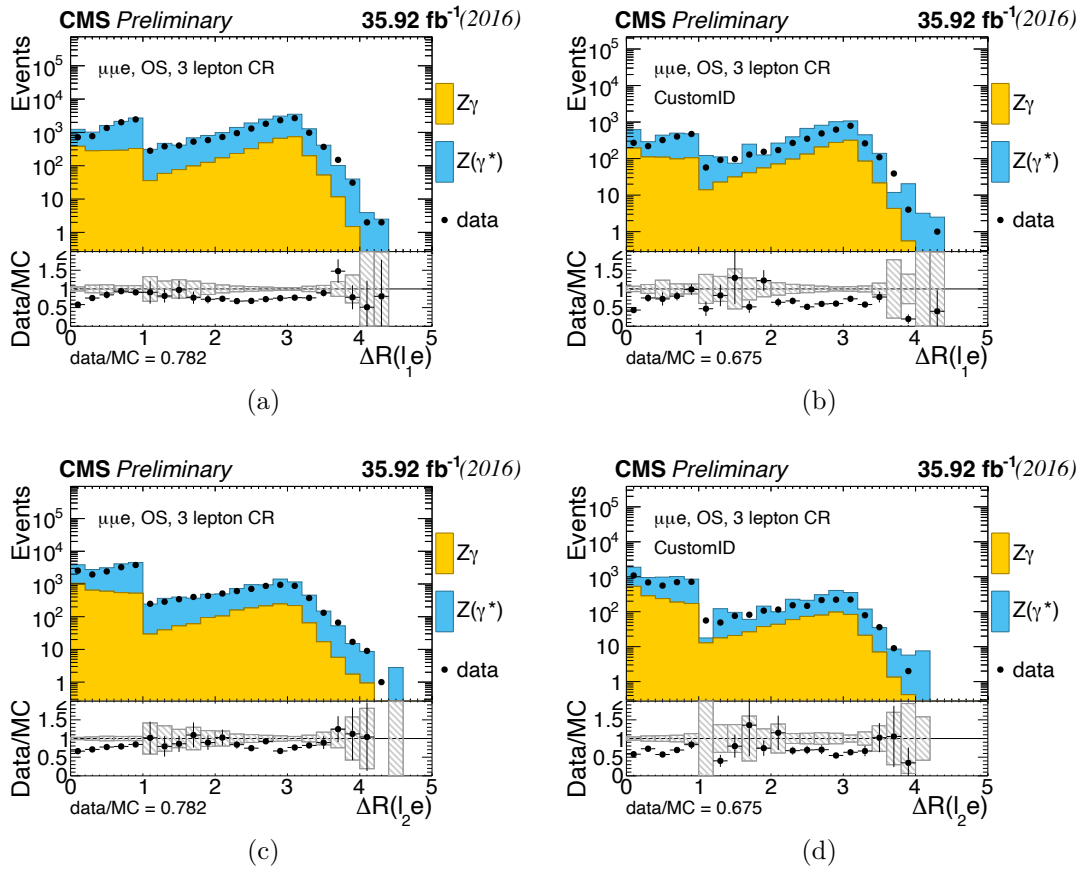


Figure E.5: Data to MC agreement plots in the 3-lepton CR for $\Delta R(\ell_1 e)$ and $\Delta R(\ell_2 e)$. Plots on the left do not have custom electron ID applied whereas those on the right do.

E.2 Displaced electron ID efficiencies

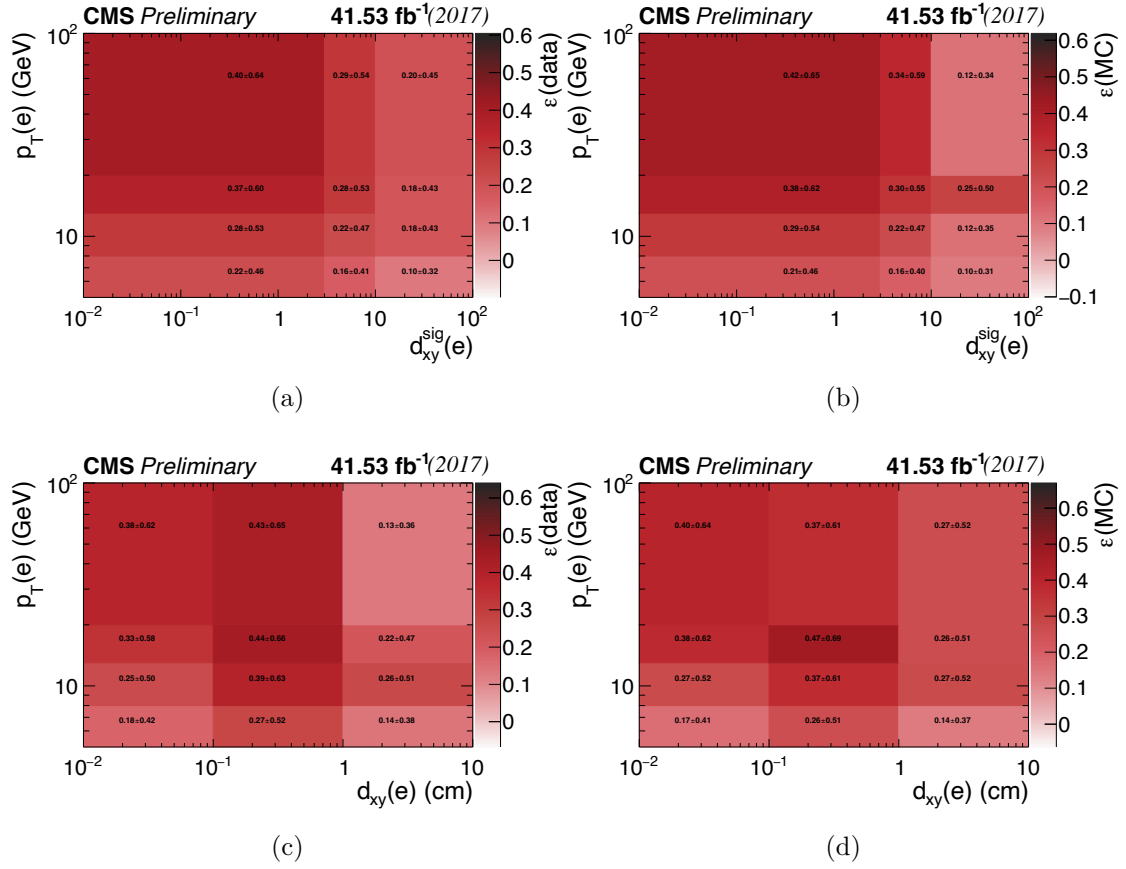


Figure E.6: 2D efficiency distributions evaluated for displaced electron identification using the raw method on both data (left) and MC (right) for kinematic variable pairs of interest (2017).

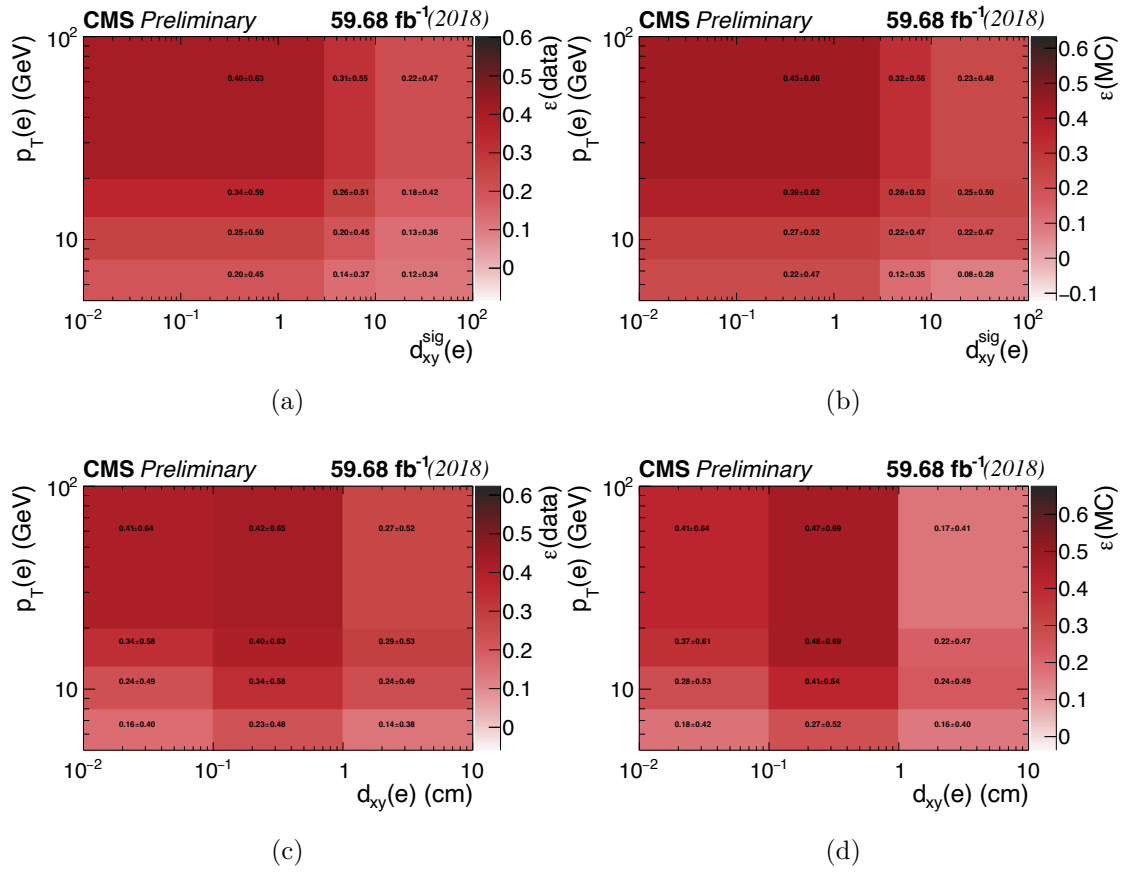


Figure E.7: 2D efficiency distributions evaluated for displaced electron identification using the raw method on both data (left) and MC (right) for kinematic variable pairs of interest (2018).

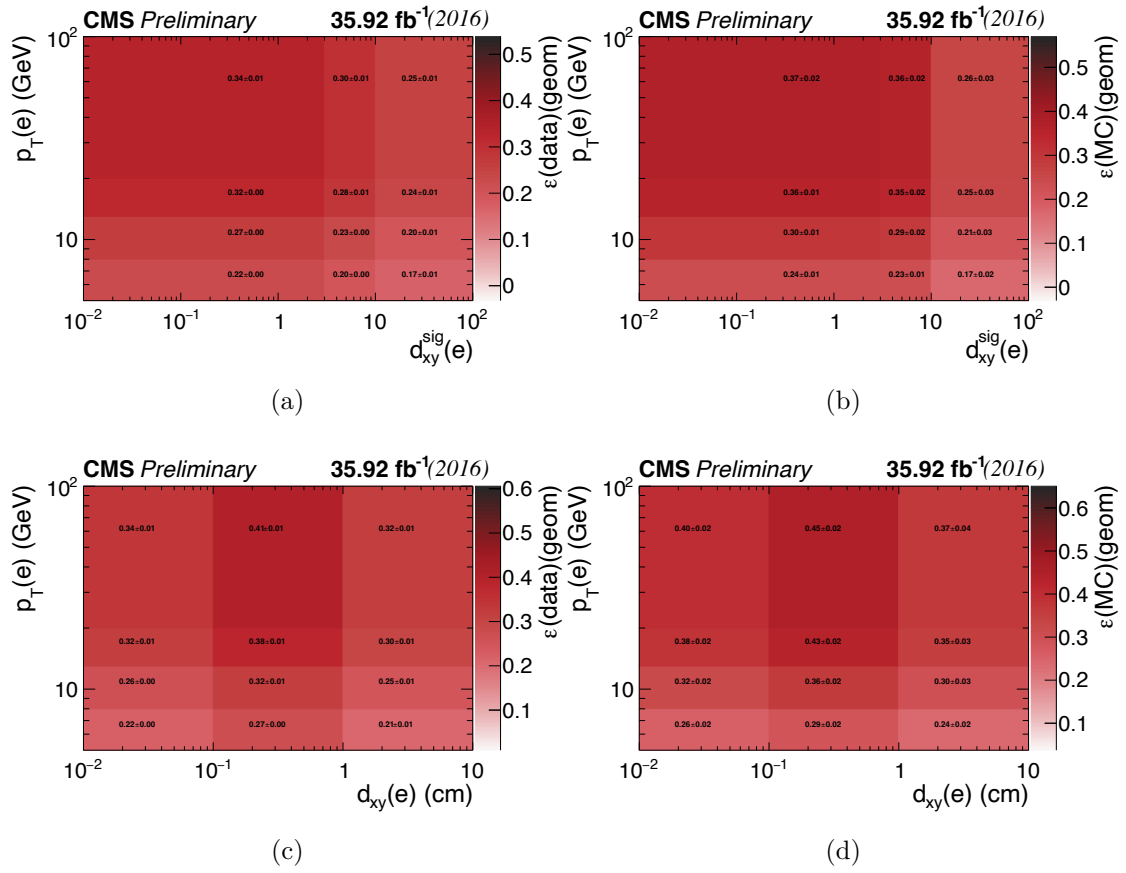


Figure E.8: 2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for kinematic variable pairs of interest (2016).

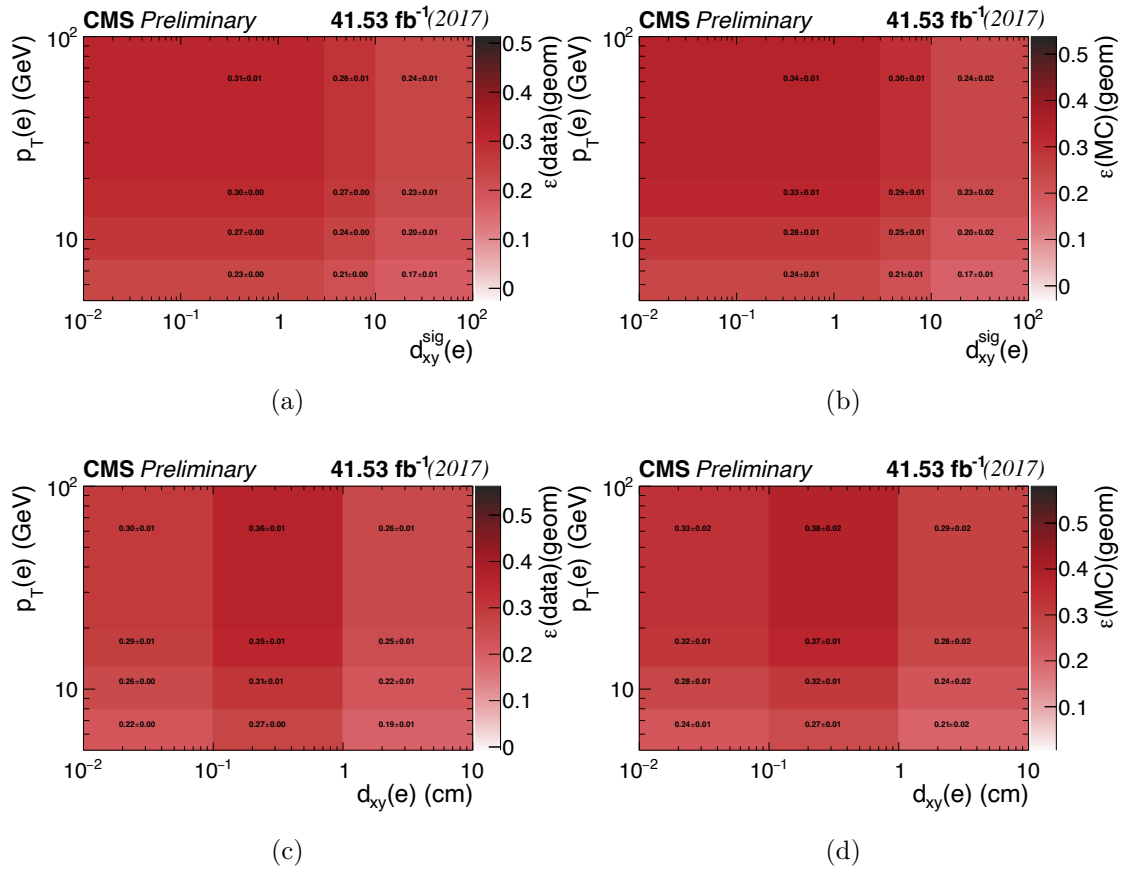


Figure E.9: 2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for kinematic variable pairs of interest (2017).

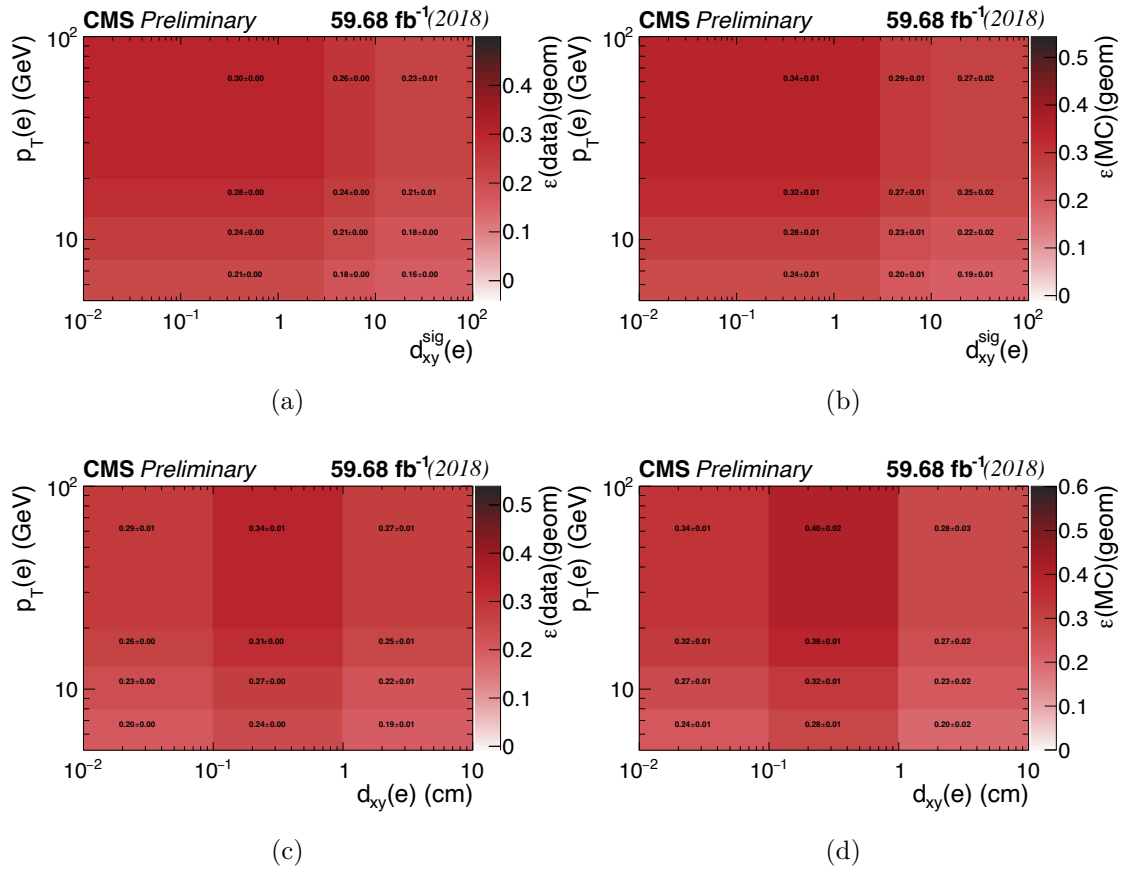
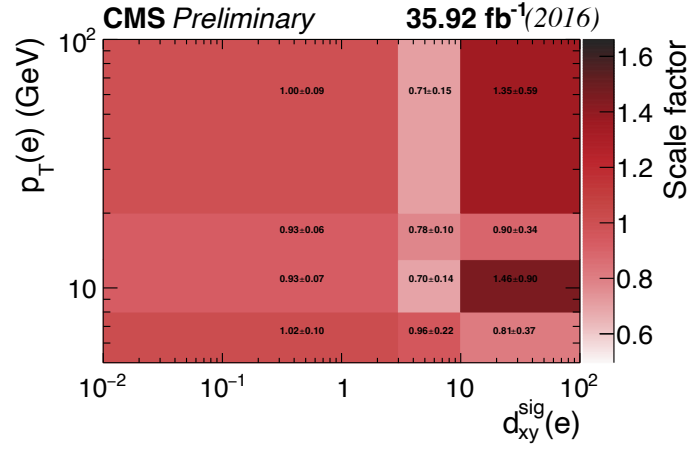
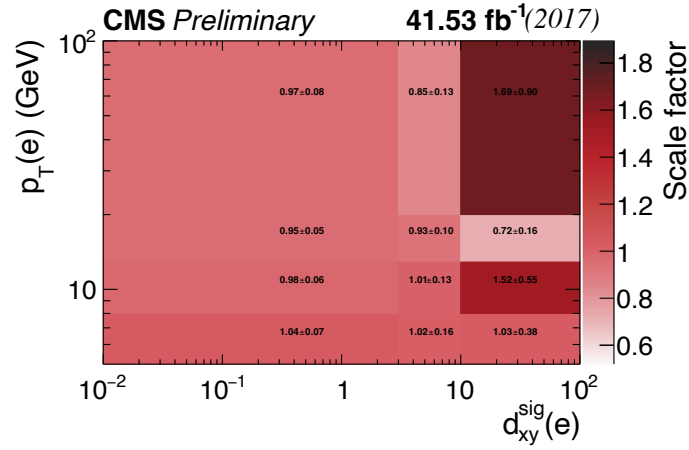


Figure E.10: 2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for kinematic variable pairs of interest (2018).

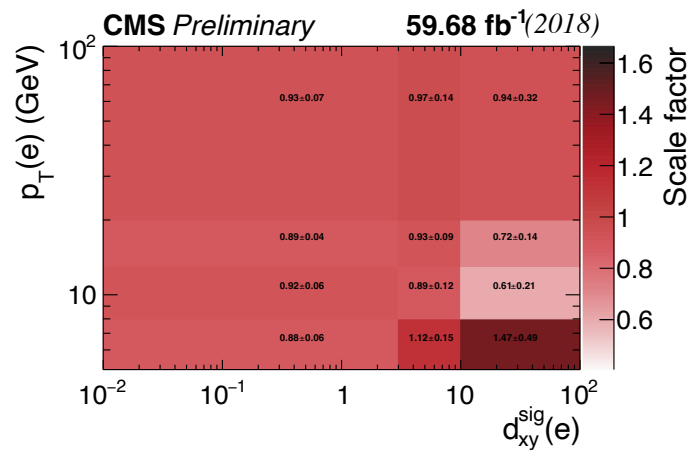
E.3 Displaced electron ID scale factors



(a)

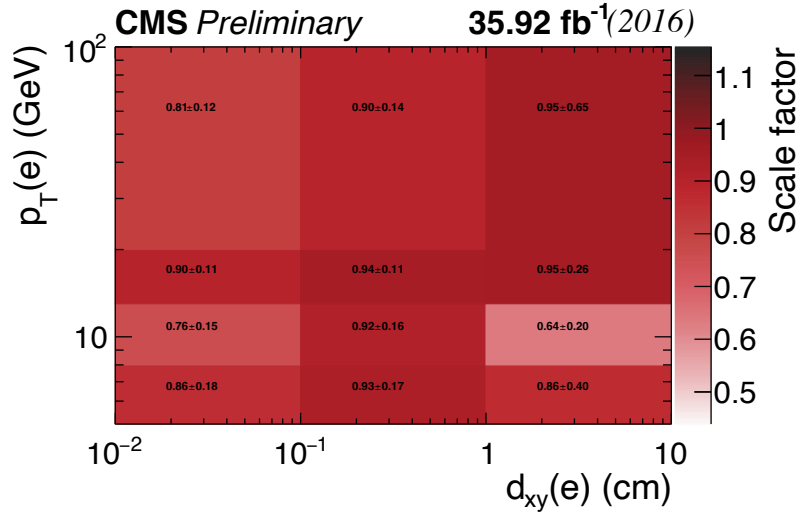


(b)

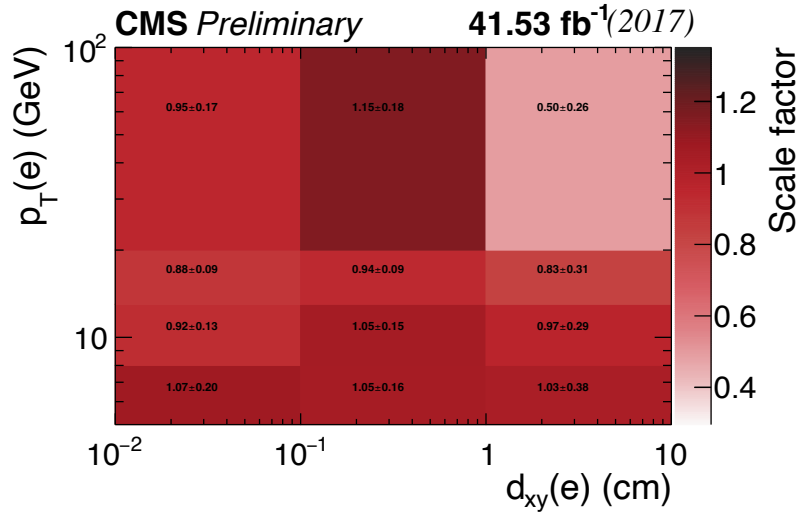


(c)

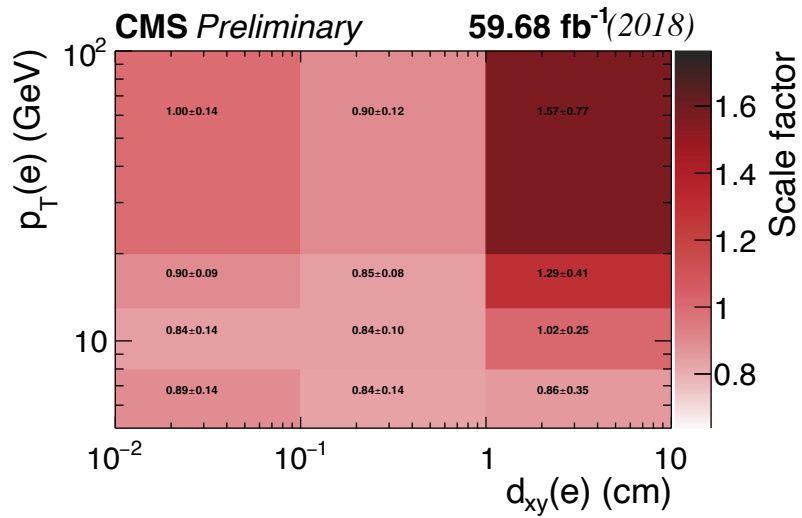
Figure E.11: Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years.



(a)

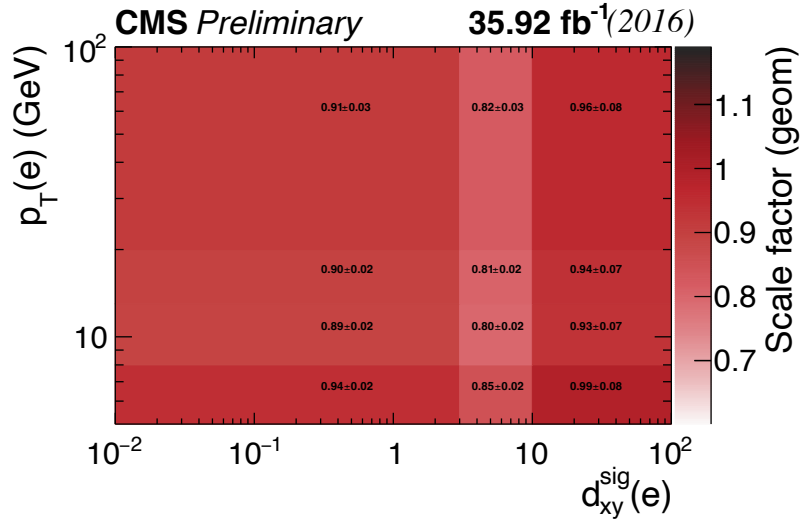


(b)

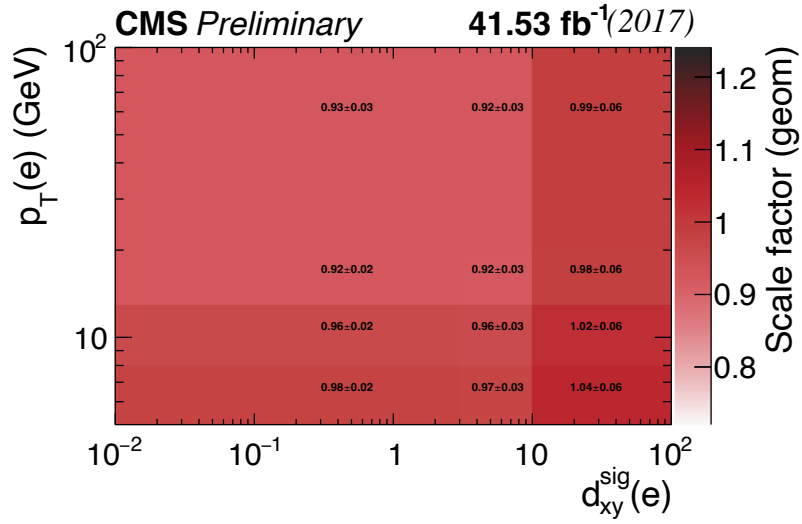


(c)

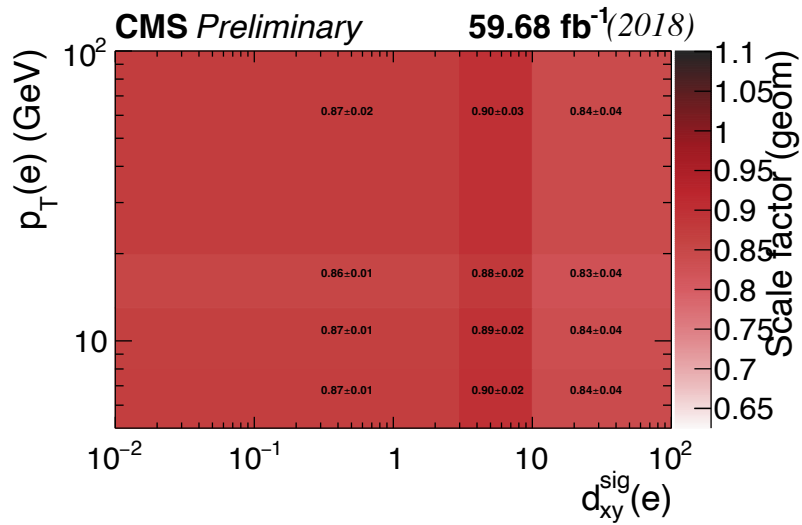
Figure E.12: Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}(e)$ kinematic variable pair for all three data taking years.



(a)

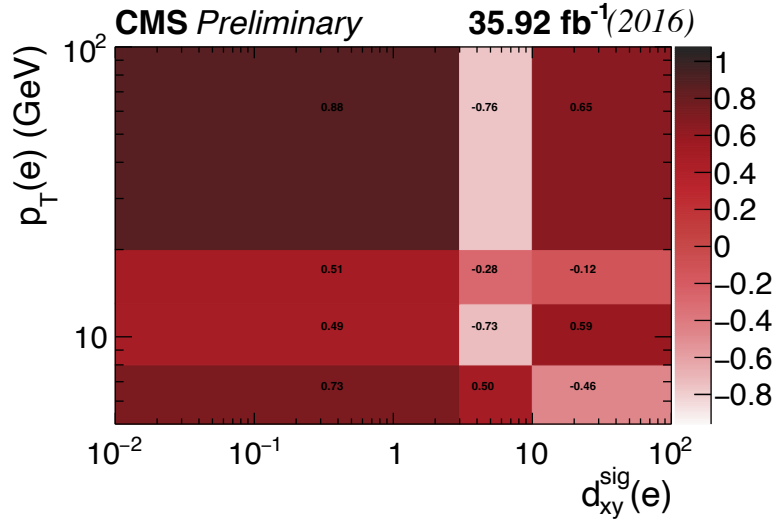


(b)

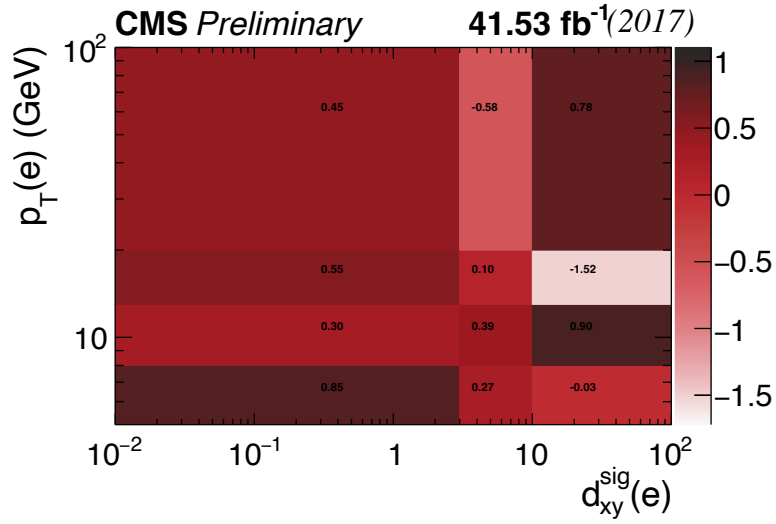


(c)

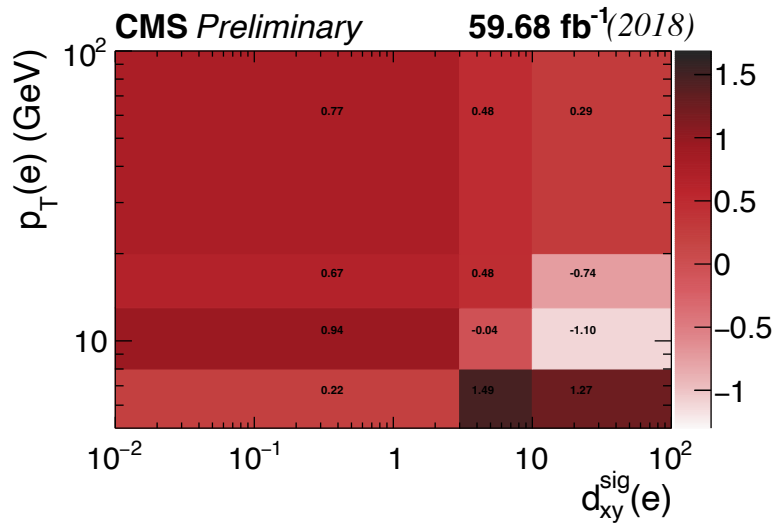
Figure E.13: Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years. Uncertainties presented are statistical only.



(a)

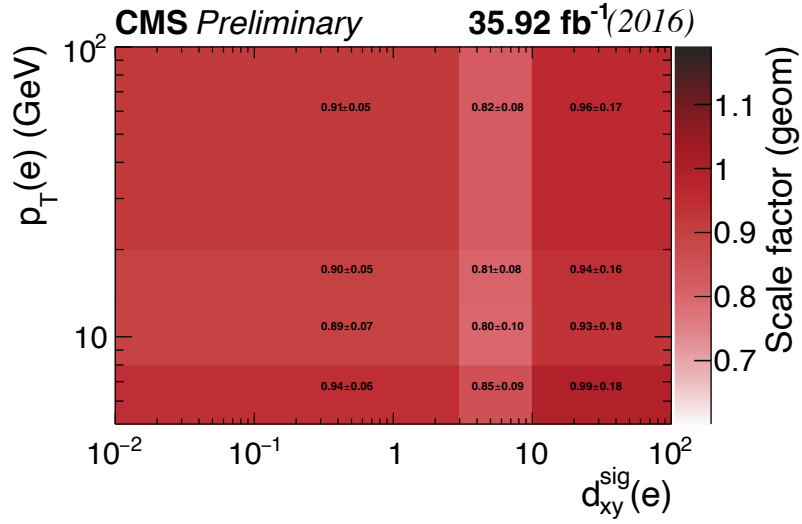


(b)

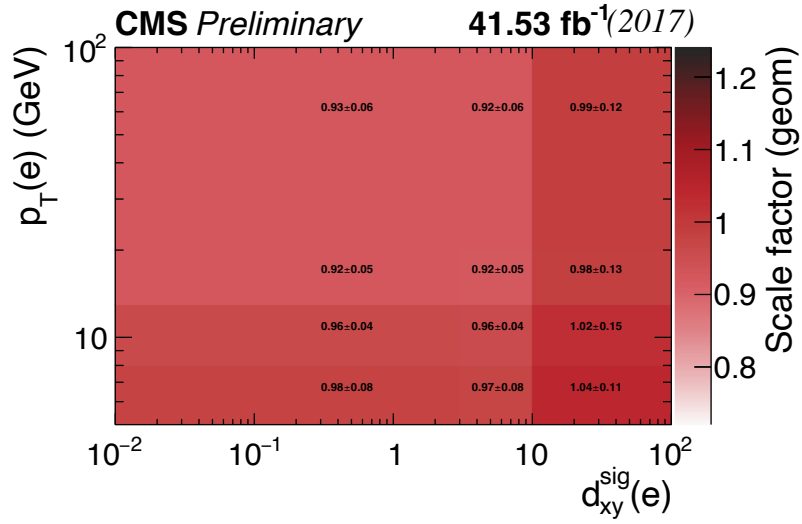


(c)

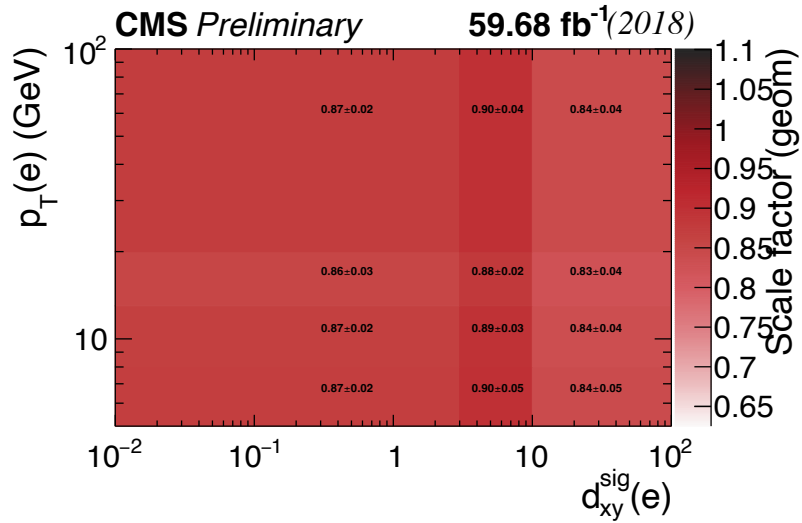
Figure E.14: Pull values estimated between the “raw” and “geometric mean” scale factor approaches on the $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years.



(a)



(b)



(c)

Figure E.15: Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years including systematic uncertainties. These scale factors are eventually propagated further in the HNL analysis.

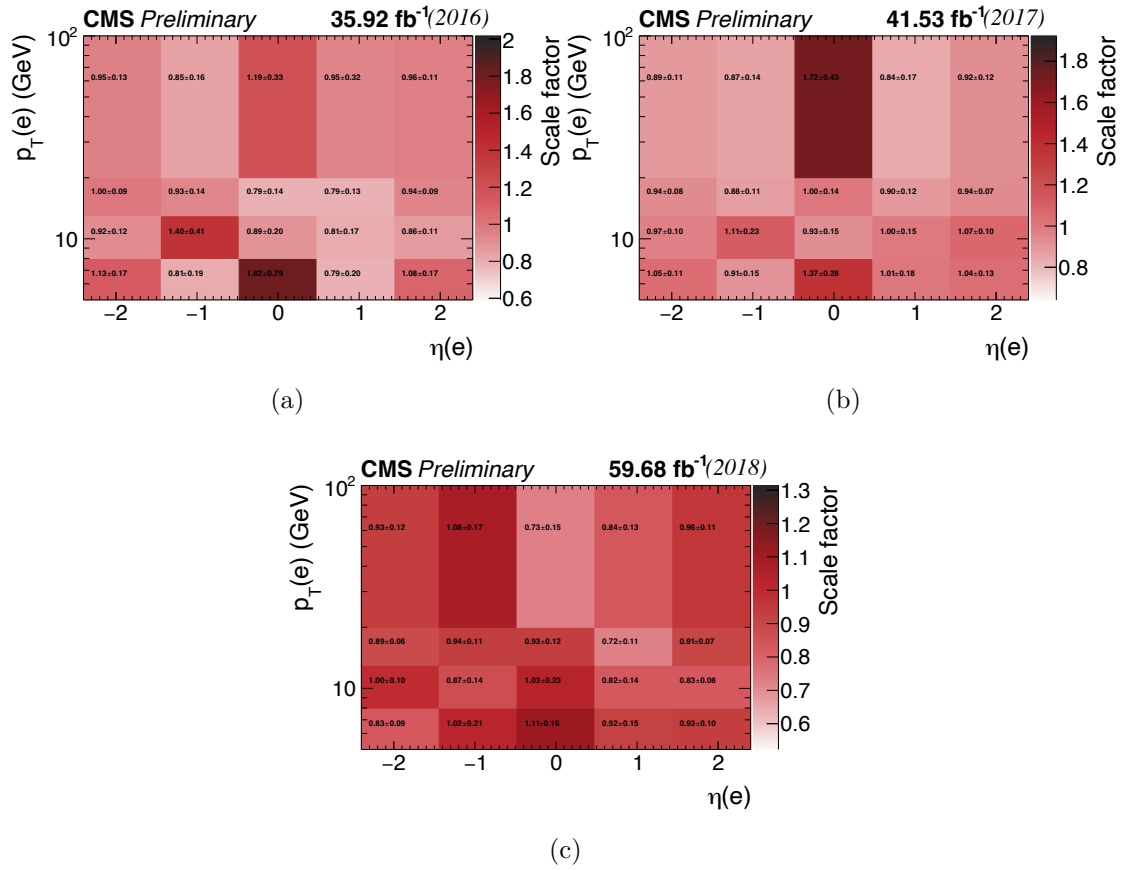


Figure E.16: Scale factors evaluated via the “raw” approach on displaced electron identification for $p_T(e)$ vs $\eta(e)$ kinematic variable pair for all three data taking years.

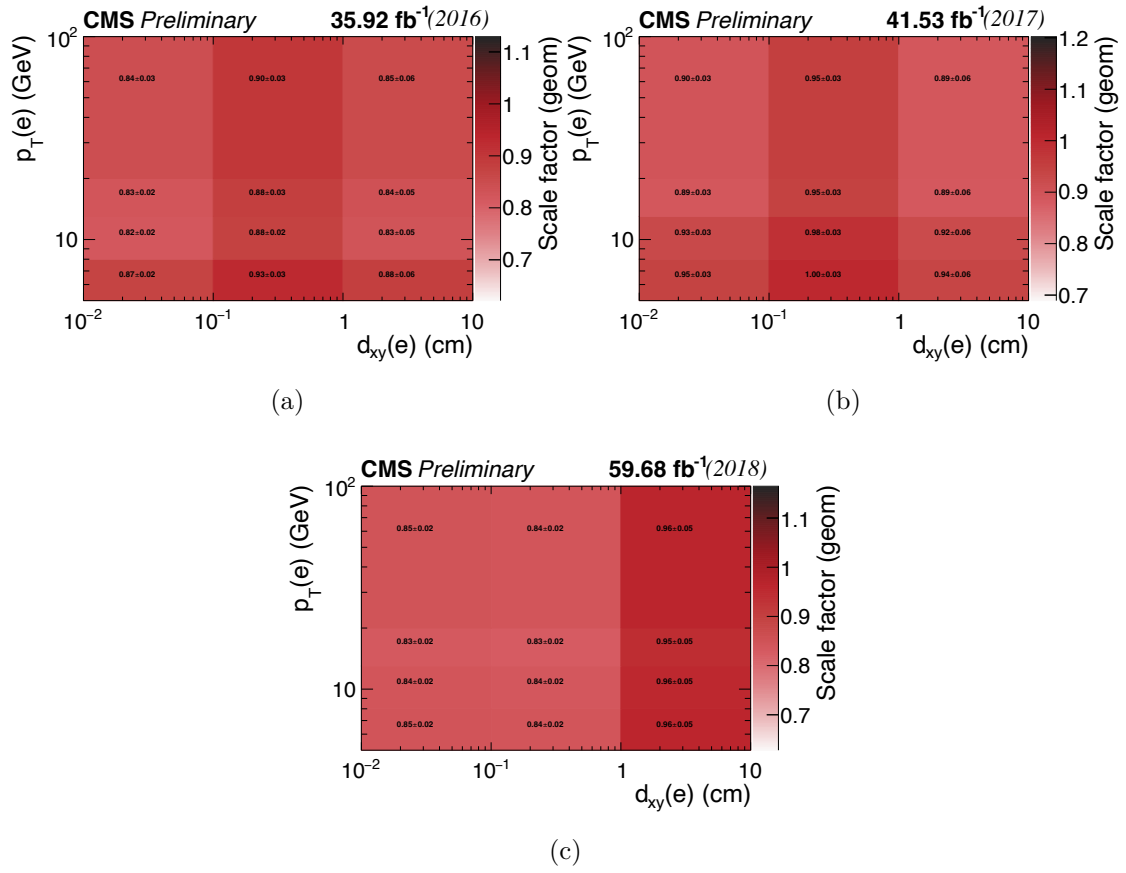


Figure E.17: Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}(e)$ kinematic variable pair for all three data taking years.

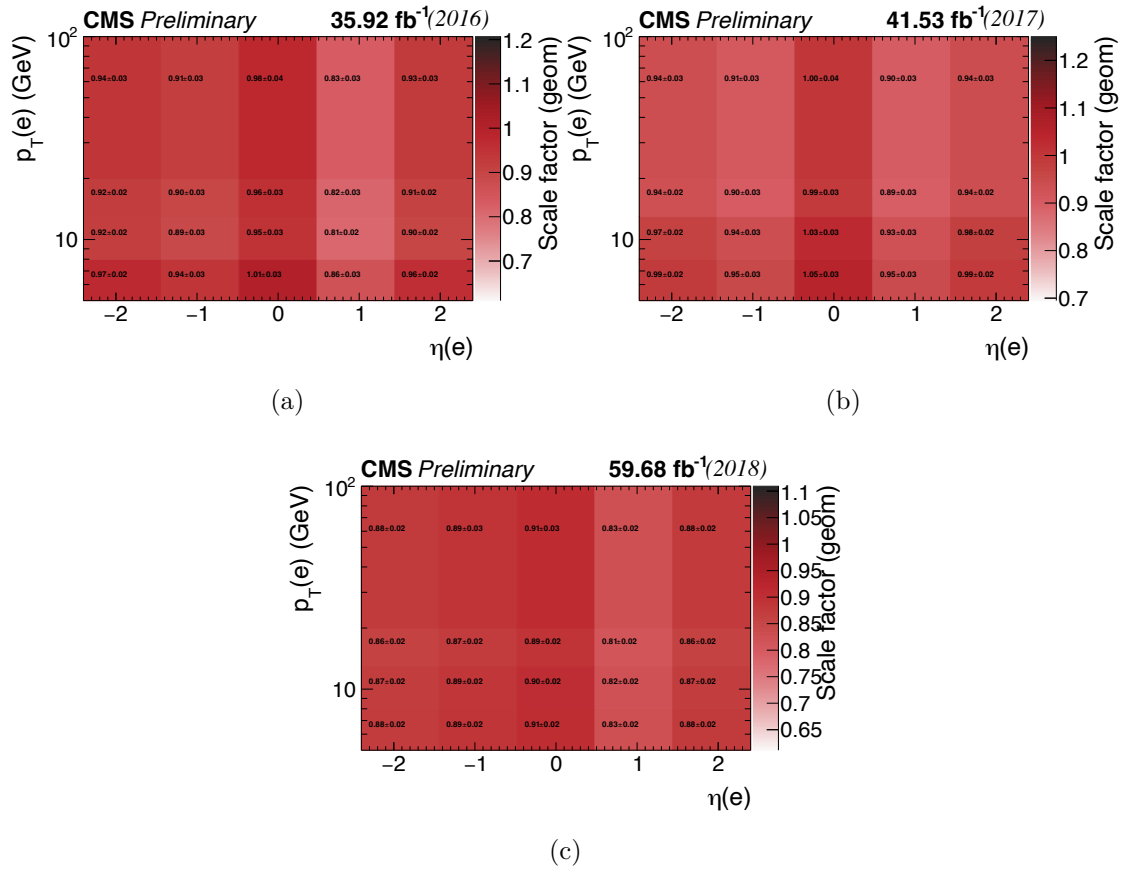


Figure E.18: Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $\eta(e)$ kinematic variable pair for all three data taking years.

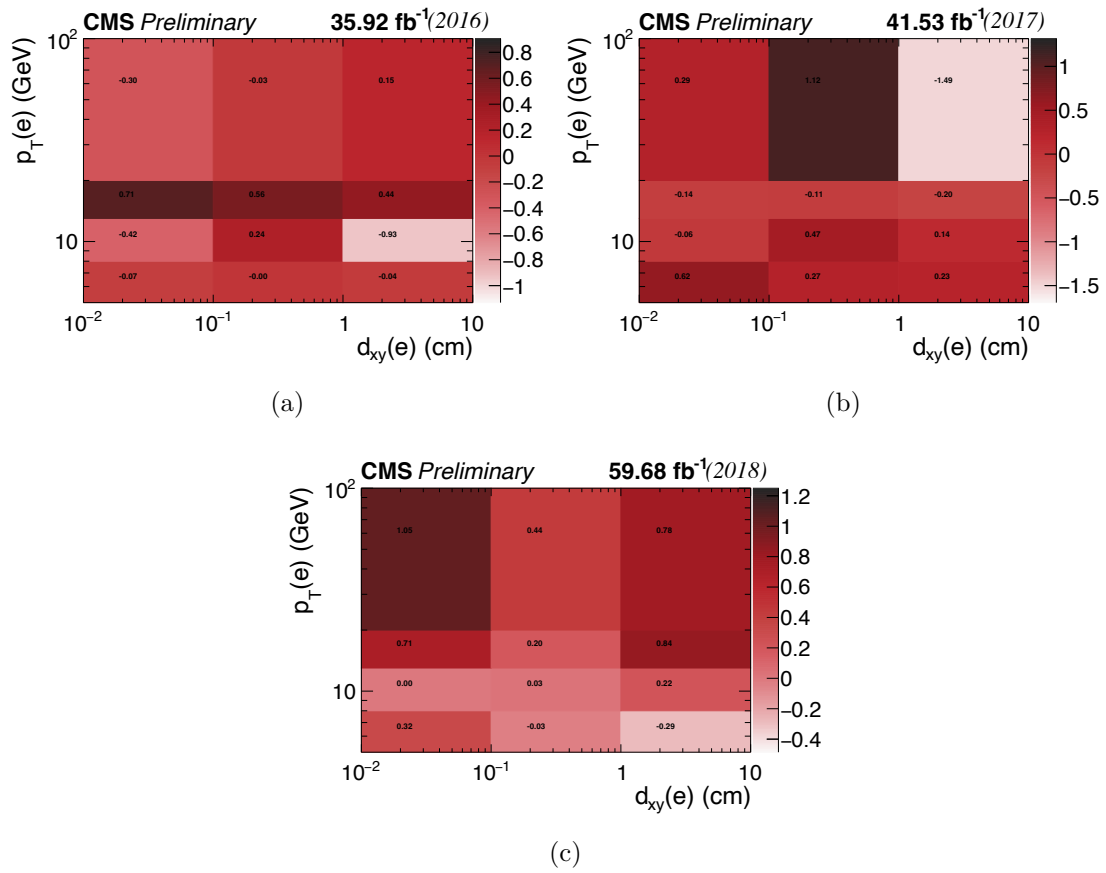


Figure E.19: Pull values estimated between the “raw” and “geometric mean” scale factor approaches on the $p_T(e)$ vs $d_{xy}(e)$ kinematic variable pair for all three data taking years.

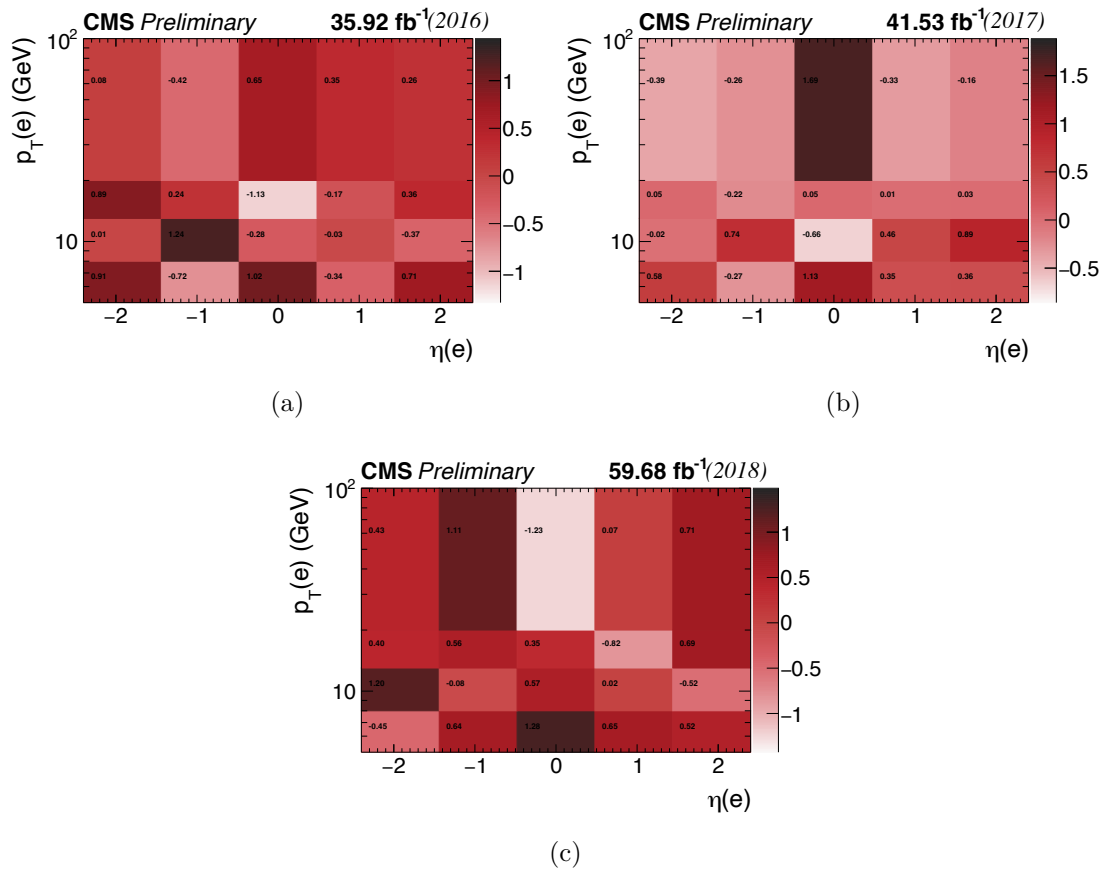


Figure E.20: Pull values estimated between the “raw” and “geometric mean” scale factor approaches on the $p_T(e)$ vs $\eta(e)$ kinematic variable pair for all three data taking years.

E.4 Displaced electron ID scale factor systematic uncertainty study

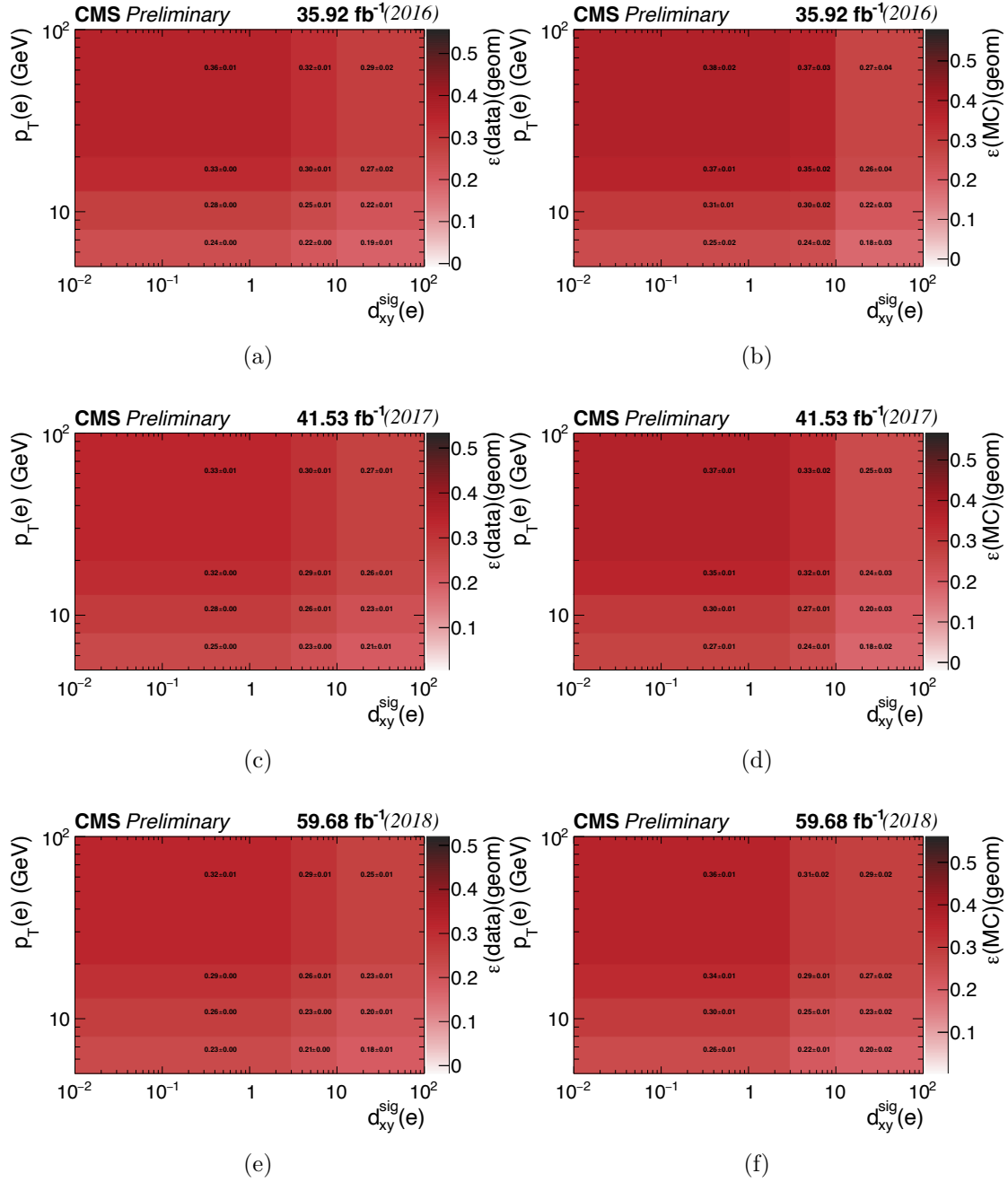


Figure E.21: 2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years in the low pile-up region.

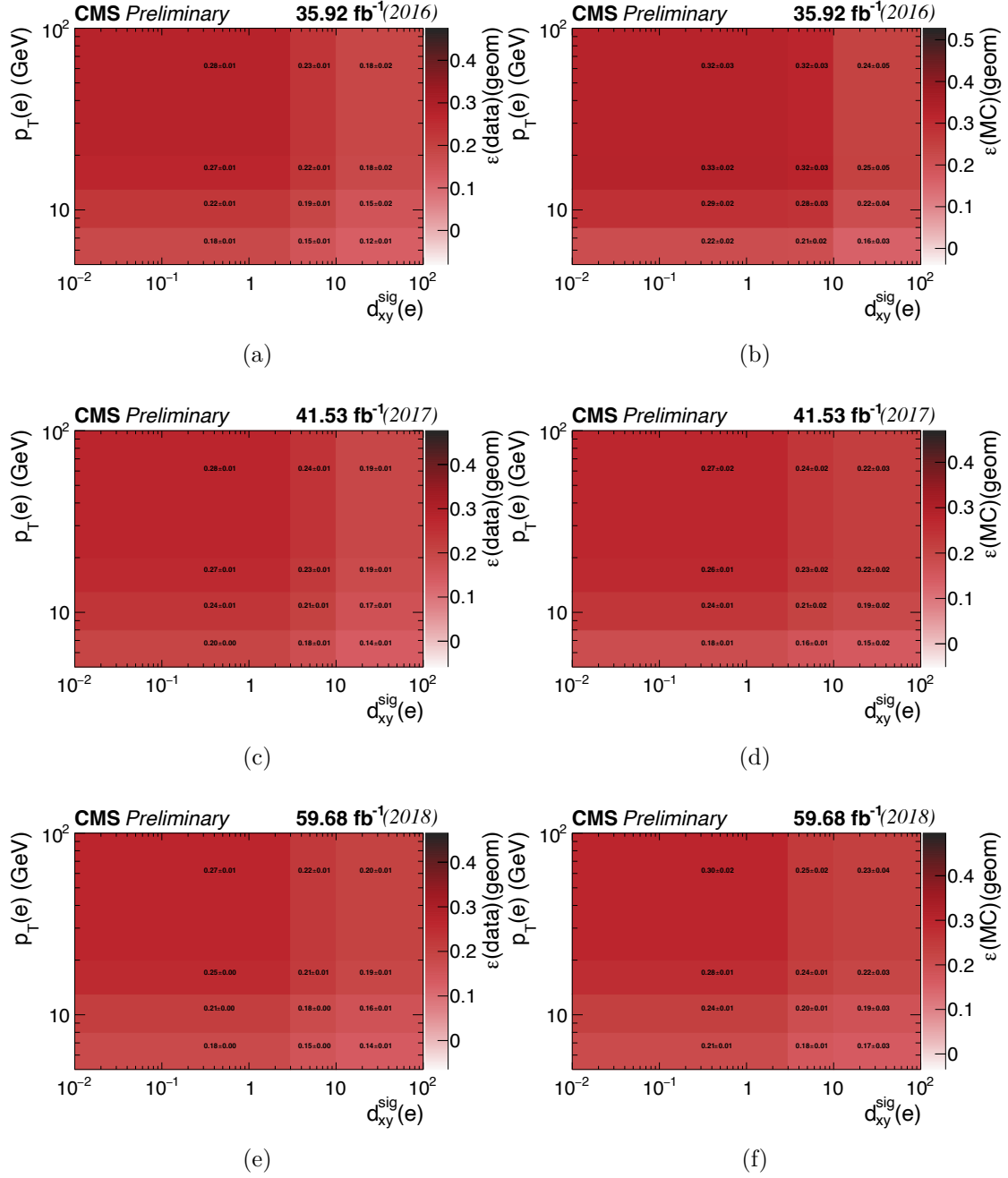


Figure E.22: 2D efficiency distributions evaluated for displaced electron identification using the geometric mean method on both data (left) and MC (right) for $p_T(e)$ vs $d_{xy}^{\text{sig}}(e)$ kinematic variable pair for all three data taking years in the high pile-up region.

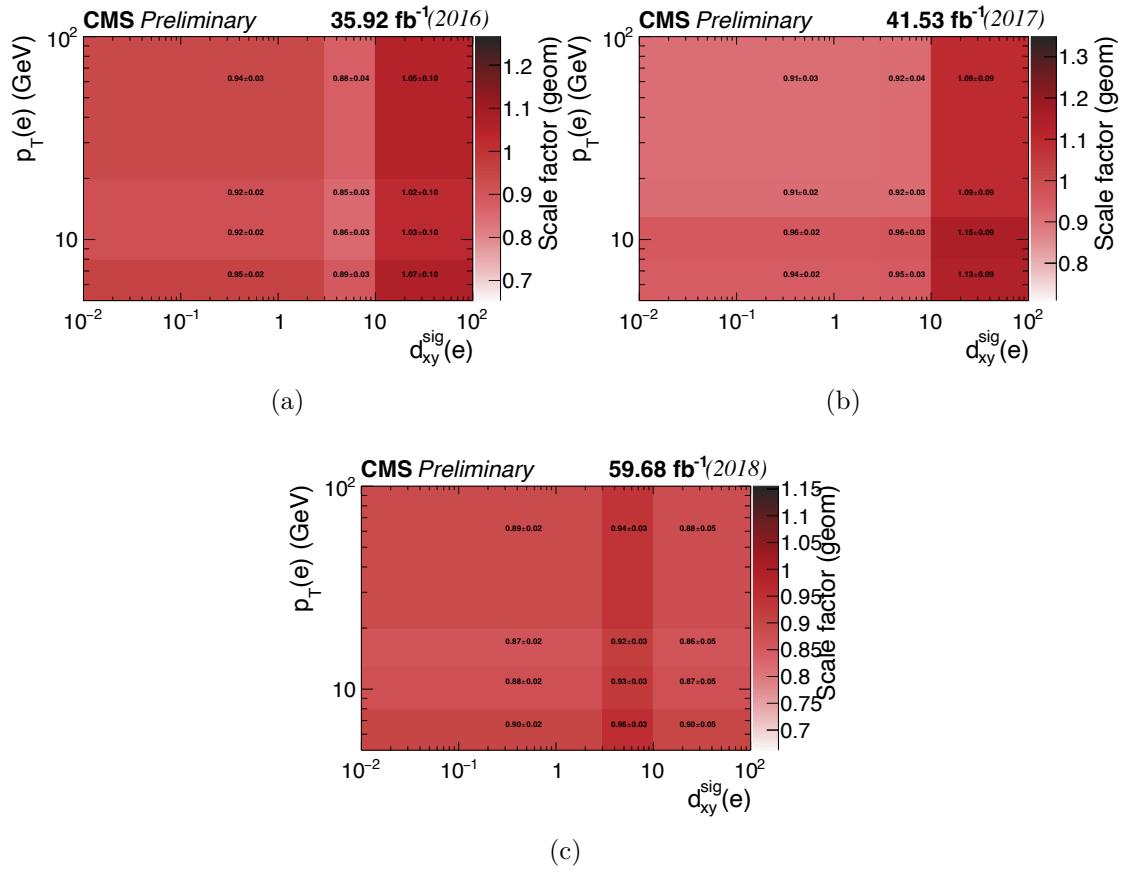


Figure E.23: Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years in the low pile-up region.

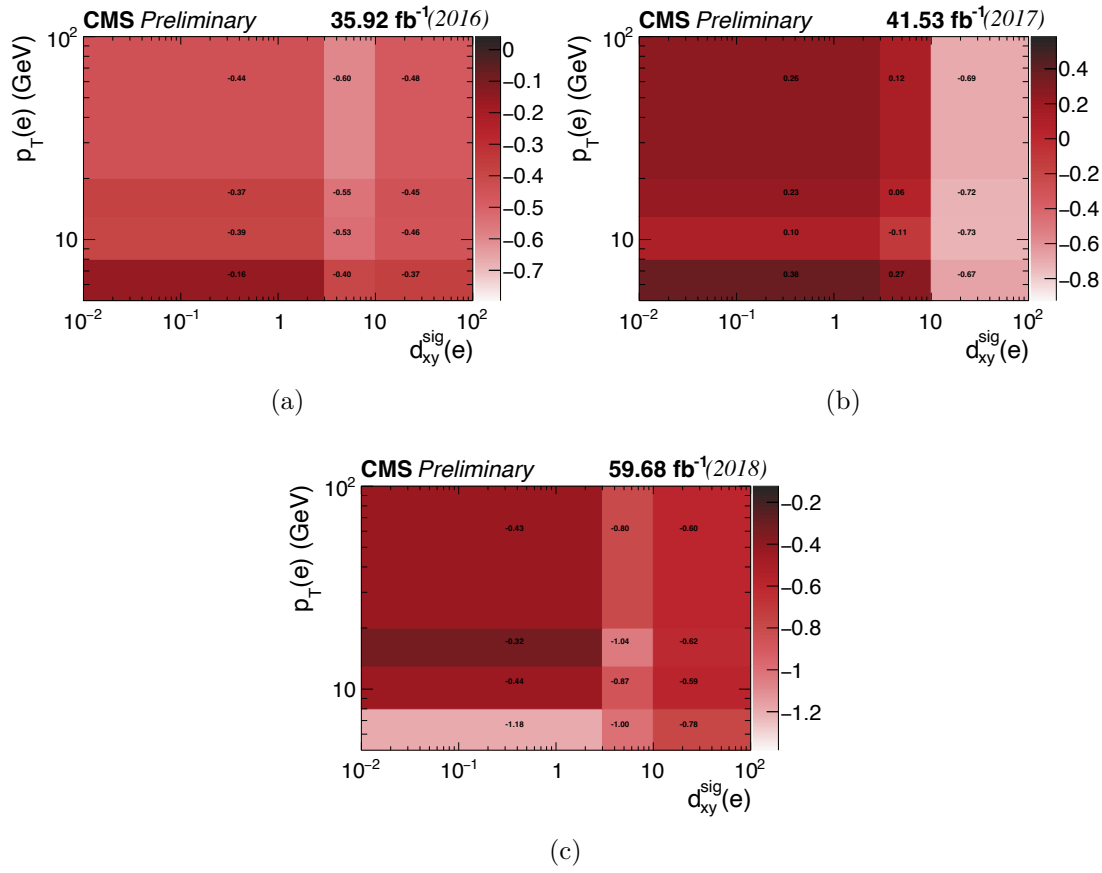


Figure E.24: Pull values estimated between the inclusive scale factors and those evaluated in the low pile-up region on the $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years.

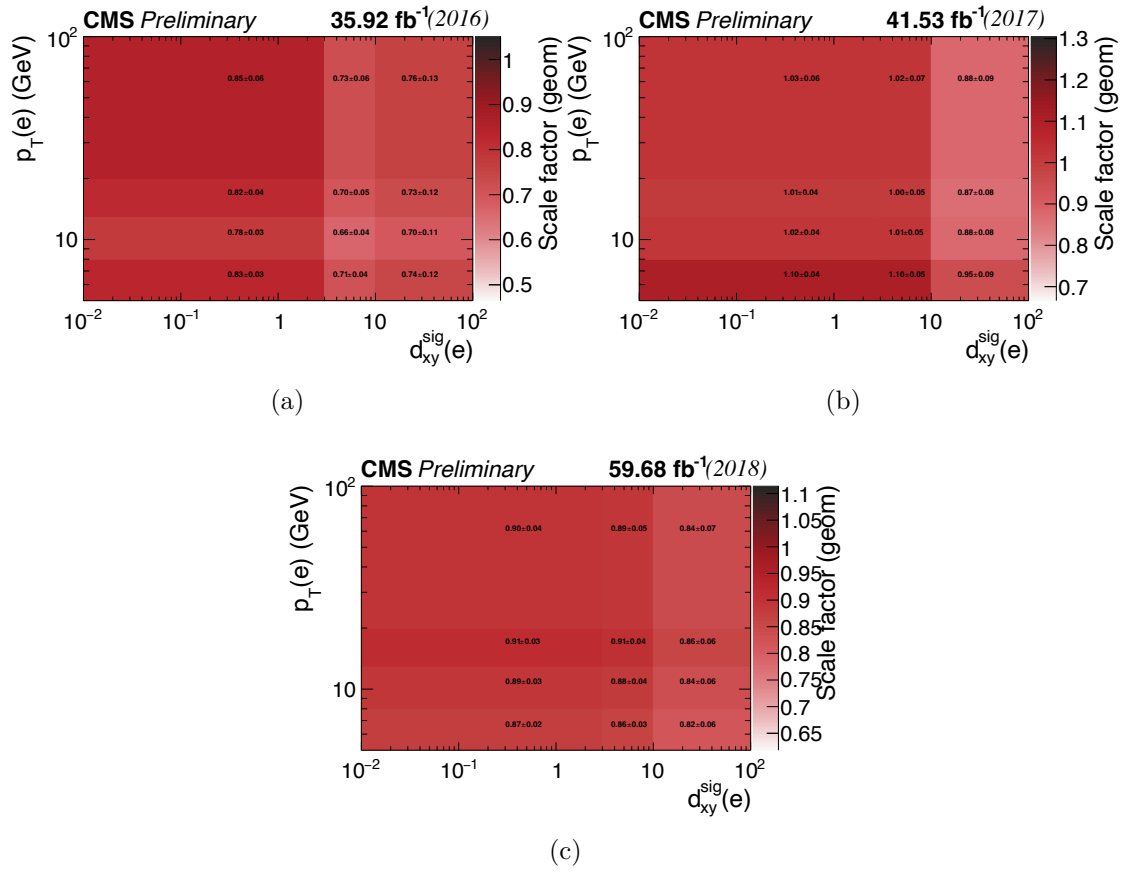


Figure E.25: Scale factors evaluated via the “geometric mean” approach on displaced electron identification for $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years in the high pile-up region.

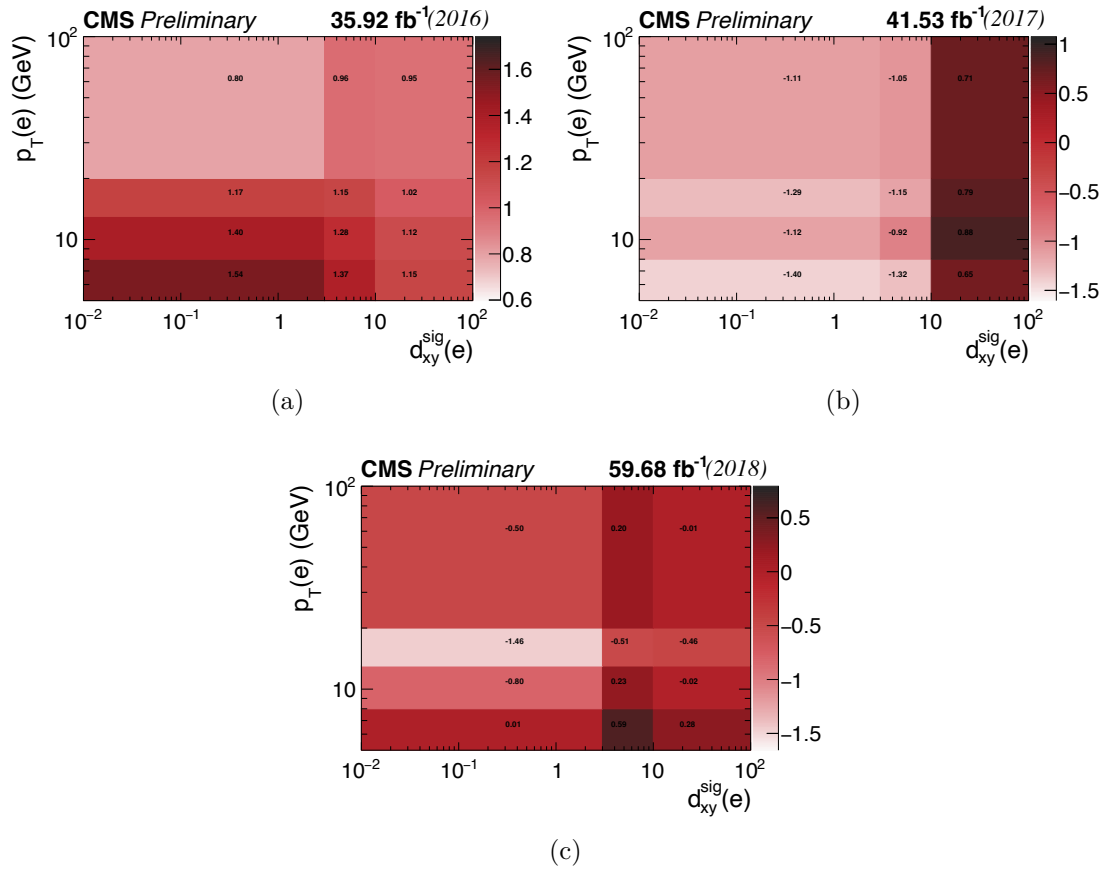


Figure E.26: Pull values estimated between the inclusive scale factors and those evaluated in the high pile-up region on the $p_T(e)$ vs $d_{xy}^{sig}(e)$ kinematic variable pair for all three data taking years.

Appendix F

Complementary material to the BDT event classifier study

F.1 2D heatmaps between the tagger score and BDT input features

Below are presented an extensive set of heatmap distributions of background processes between the displaced jet tagger information and various BDT input features considered, divided in dilepton flavour subcategories. The mutual information coefficient is shown at the bottom of each plot and the distributions are subsampled according to Sec. 8.3.

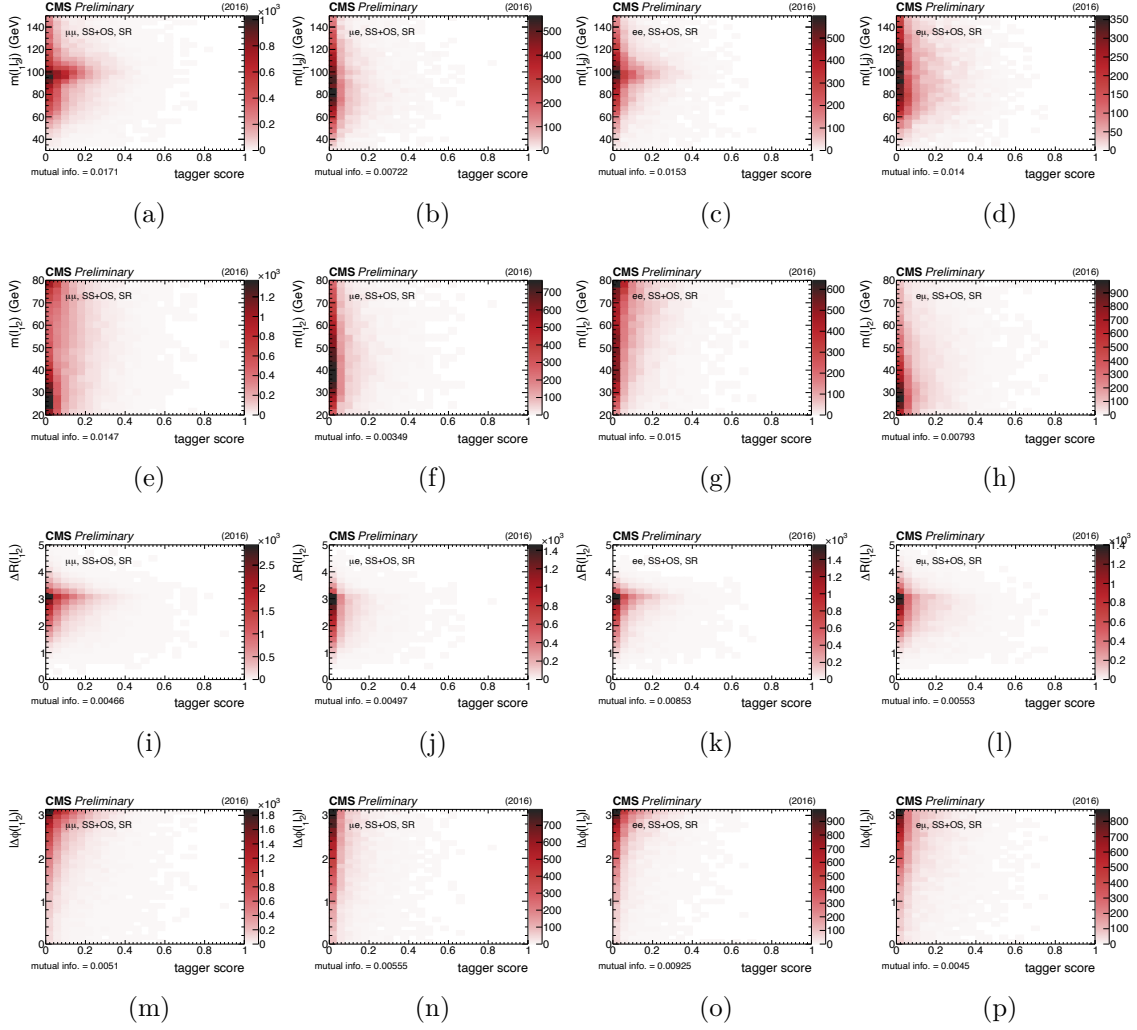


Figure F.1: 2D heatmaps between the tagger score and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).

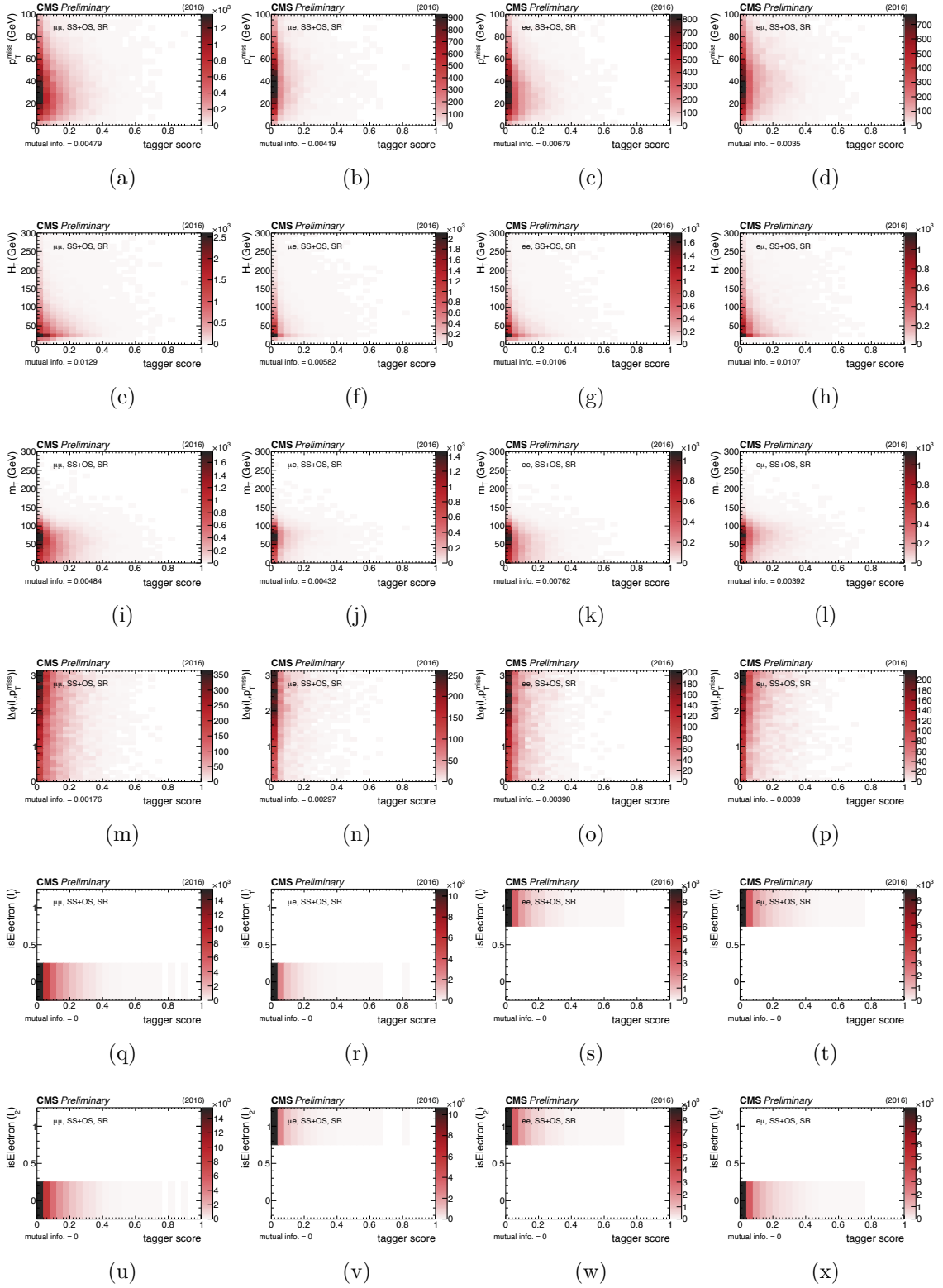


Figure F.2: 2D heatmaps between the tagger score and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).

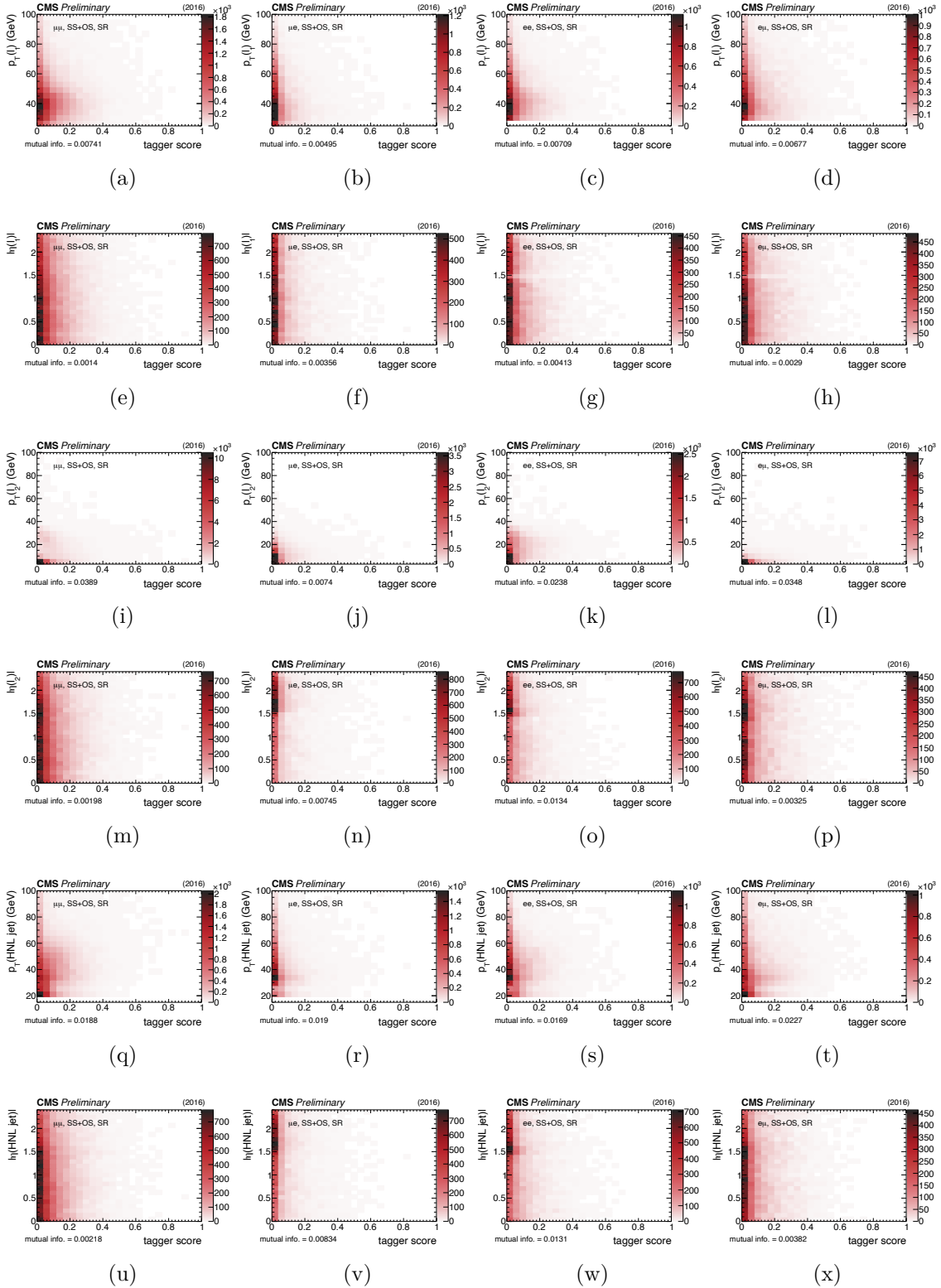


Figure F.3: 2D heatmaps between the tagger score and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined ($SS+OS$).

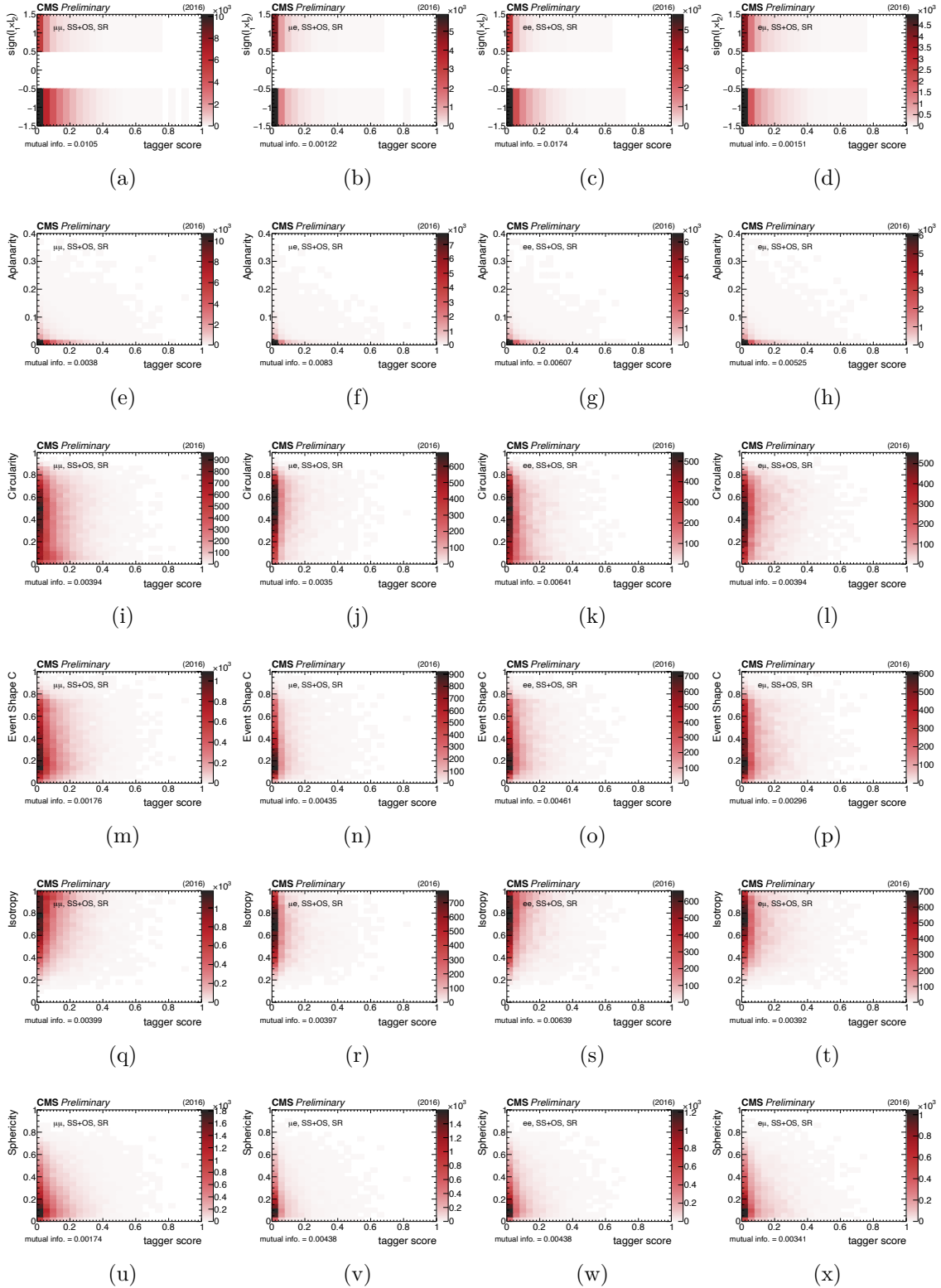


Figure F.4: 2D heatmaps between the tagger score and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).

F.2 2D heatmaps between the $m(\ell_1, \ell_2, j^*)$ variable and BDT input features

Below are presented an extensive set of heatmap distributions of background processes between the $m(\ell_1, \ell_2, j^*)$ variable and various BDT input features considered, divided in dilepton flavour subcategories. The mutual information coefficient is shown at the bottom of each plot and the distributions are subsampled according to Section 8.3.

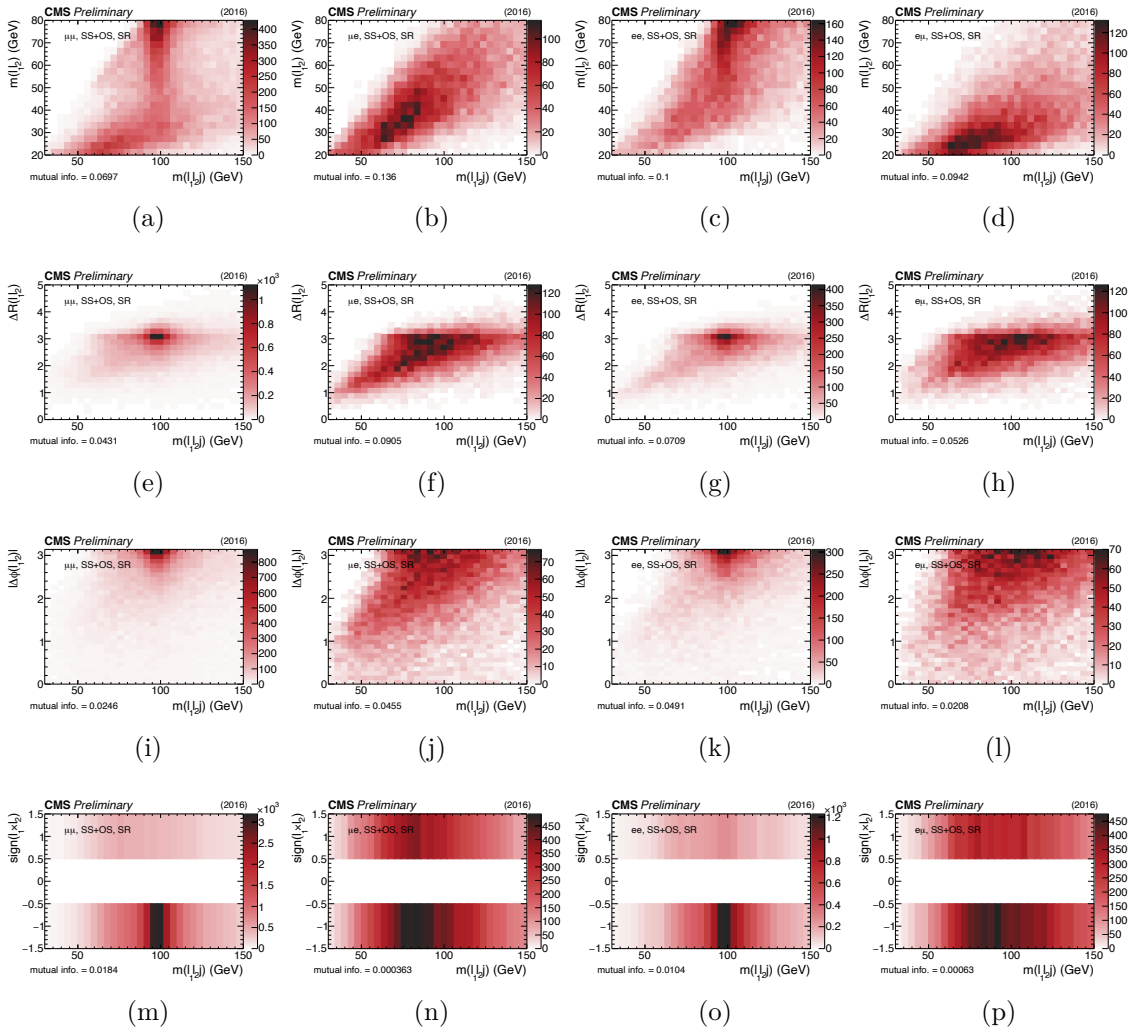


Figure F.5: 2D heatmaps between $m(\ell_1, \ell_2, j^*)$ and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).

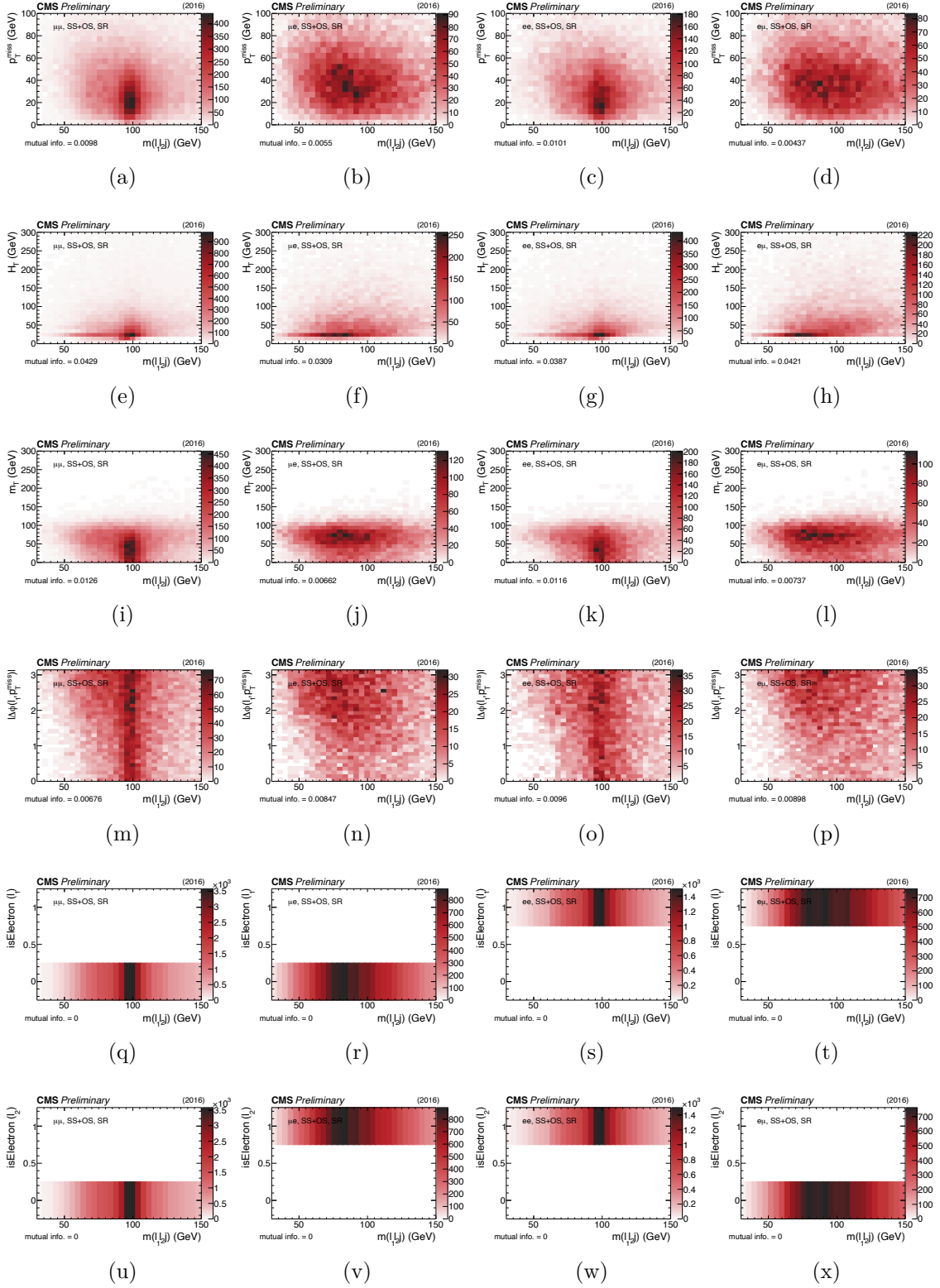


Figure F.6: 2D heatmaps between $m(\ell_1, \ell_2, j^*)$ and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).

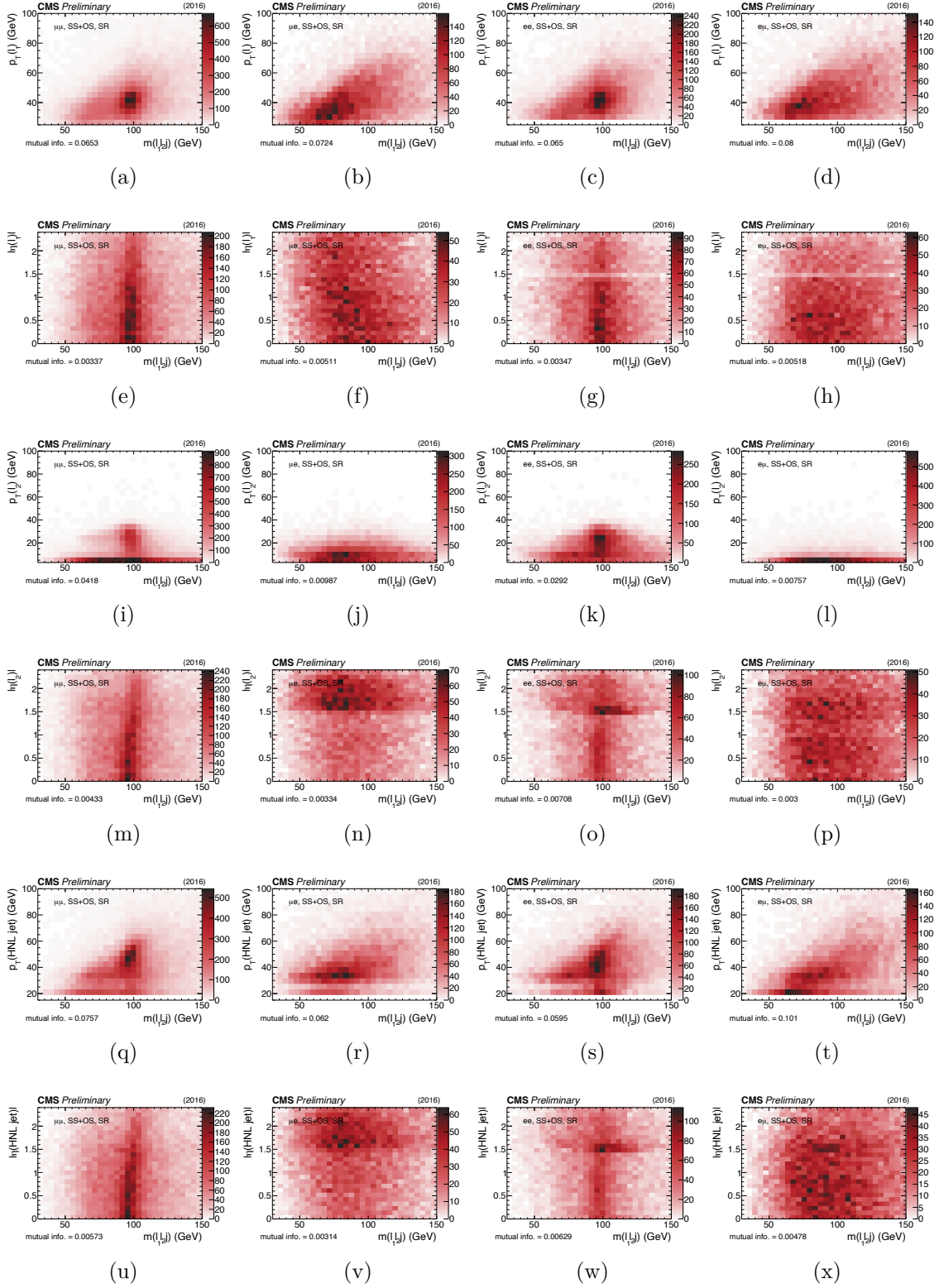


Figure F.7: 2D heatmaps between $m(\ell_1, \ell_2, j^*)$ and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).

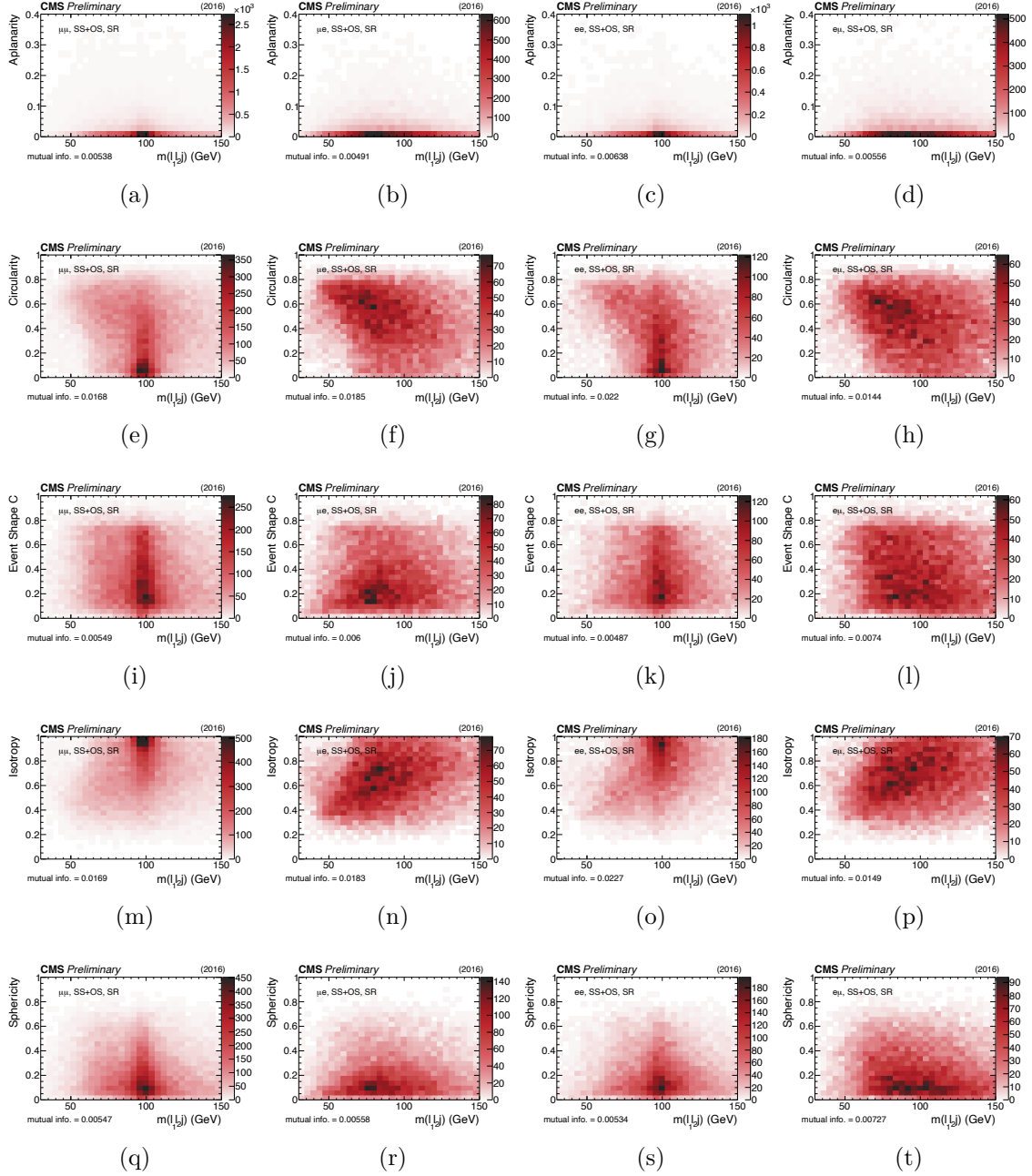


Figure F.8: 2D heatmaps between $m(\ell_1, \ell_2, j^*)$ and various BDT feature inputs in the signal region investigating potential correlations between them. The plots contain only background processes the BDT is trained on, subsampled based on their expected cross-sections (2016 MC) and are presented by the four dilepton flavour categories the BDT is trained on, with the dilepton charge combined (SS+OS).

F.3 Correlation studies between the $m(\ell_1, \ell_2, j^*)$ variable and BDT inputs

By looking at an extensive set of heatmap distributions in App. F.2 between the $m(\ell_1, \ell_2, j^*)$ variable and various BDT input features, information on the extent of correlation between the $m(\ell_1, \ell_2, j^*)$ variable and BDT input features is obtained. Stronger correlation is observed in the general shape and mutual information value between the $m(\ell_1, \ell_2, j^*)$ variable and H_T , $p_T(\ell_1)$, $p_T(j^*)$, $m(\ell_1, \ell_2)$, $\Delta R(\ell_1, \ell_2)$ and $|\Delta\phi(\ell_1, \ell_2)|$. Most of these variables are components of the $m(\ell_1, \ell_2, j^*)$ by definition. Hence, it is not surprising that correlation is identified between $m(\ell_1, \ell_2, j^*)$ and them. The driving variable of H_T is $p_T(j^*)$, which is a component of $m(\ell_1, \ell_2, j^*)$.

In Figs. F.9 & F.10, rankings of potential BDT input features by their extent of correlation with the $m(\ell_1, \ell_2, j^*)$ variable in the four dilepton flavour categories (dilepton charge combined, 2016 simulation) have been constructed using the Pearson correlation coefficient and the mutual information as quantitative metrics. Both metrics prove that the $p_T(j^*)$, dilepton mass, $p_T(\ell_1)$, H_T , $\Delta R(\ell_1, \ell_2)$ and $|\Delta\phi(\ell_1, \ell_2)|$ are the most correlated variables with $m(\ell_1, \ell_2, j^*)$, providing information on building BDT models that are uncorrelated from the $m(\ell_1, \ell_2, j^*)$ variable, as well as avoiding the usage of redundant input variables (further information in Sec. 8.4).

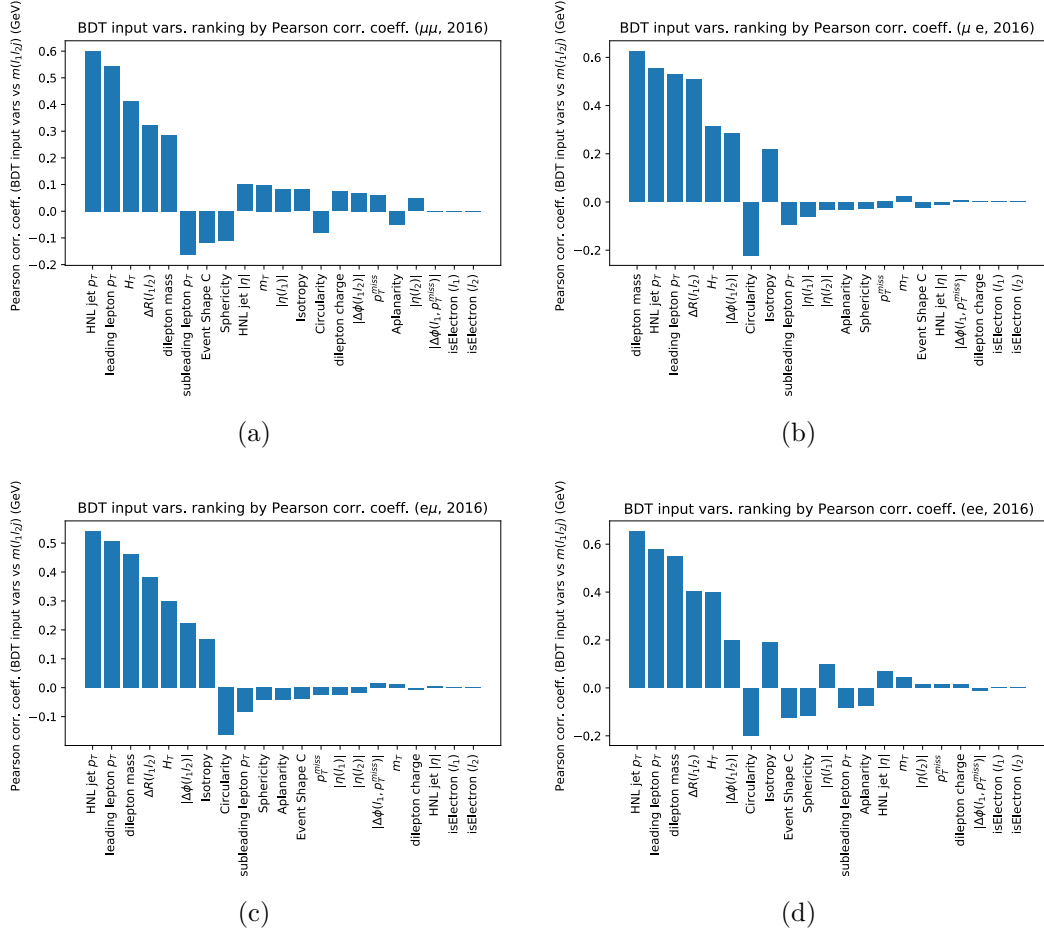


Figure F.9: BDT input feature ranking by extent of correlation to $m(\ell_1, \ell_2, j^*)$, using the Pearson correlation coefficient (2016 simulation).

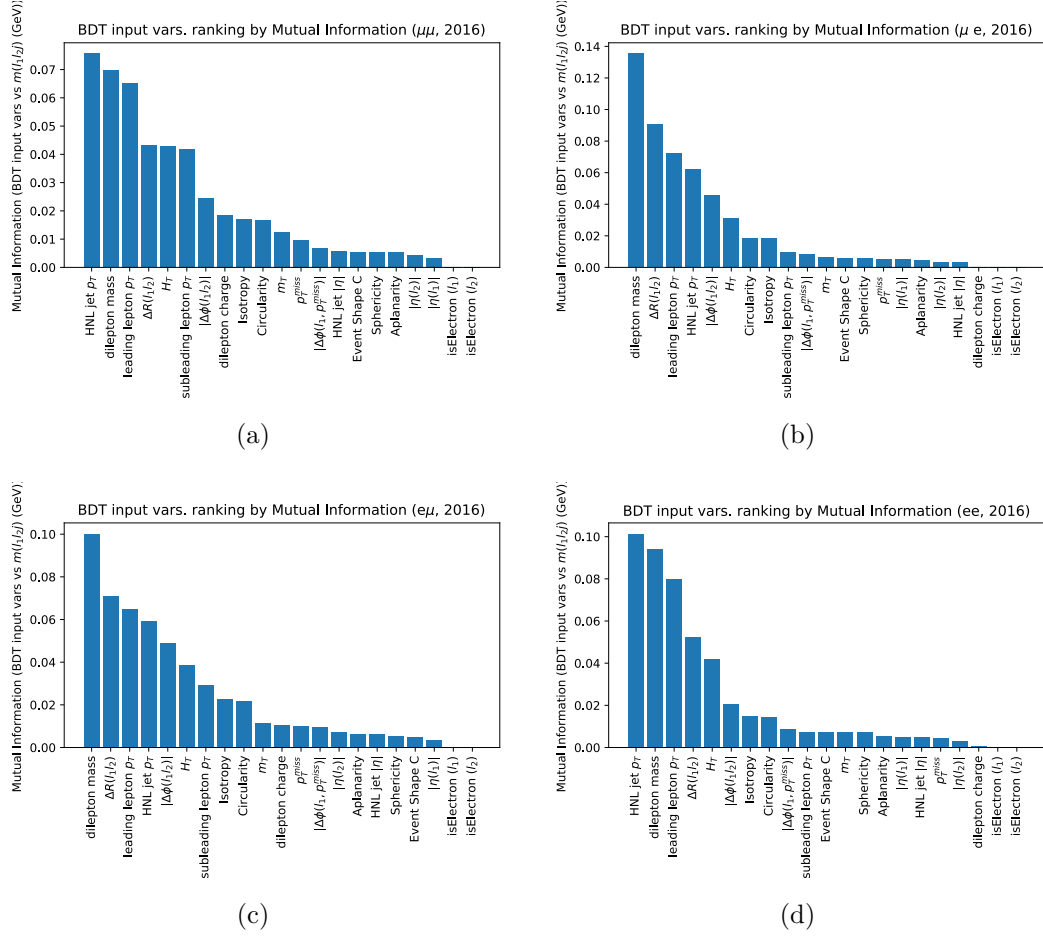


Figure F.10: BDT input feature ranking by extent of correlation to $m(\ell_1, \ell_2, j^*)$, using the mutual information metric (2016 simulation).

F.4 Performance plots of the correlated BDT model

Below are presented the performance plots of the correlated BDT model.

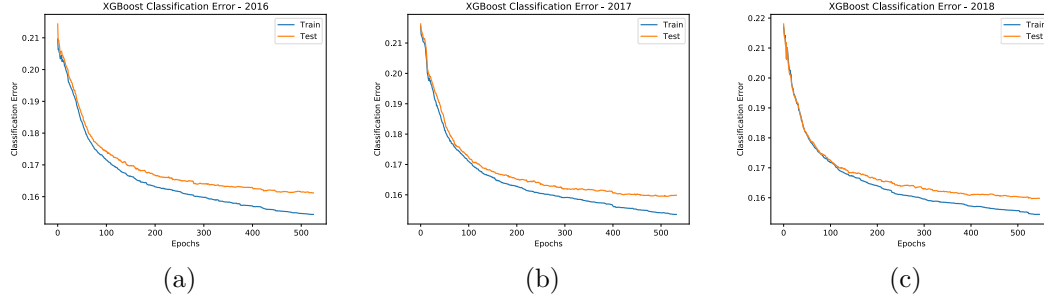


Figure F.11: Classification error development by training epoch of the correlated BDT model, as well as in the test data set.

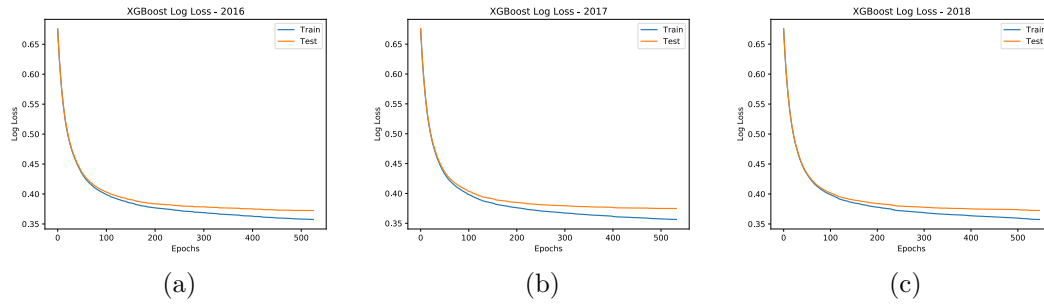


Figure F.12: Logarithmic loss development by training epoch of the correlated BDT model, as well as in the test data set.

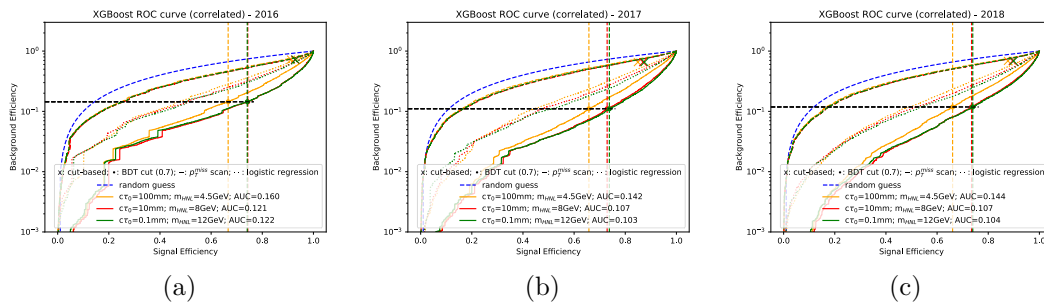


Figure F.13: BDT performance ROC curves on the evaluation data set for the correlated BDT, trained for all 3 data taking years.

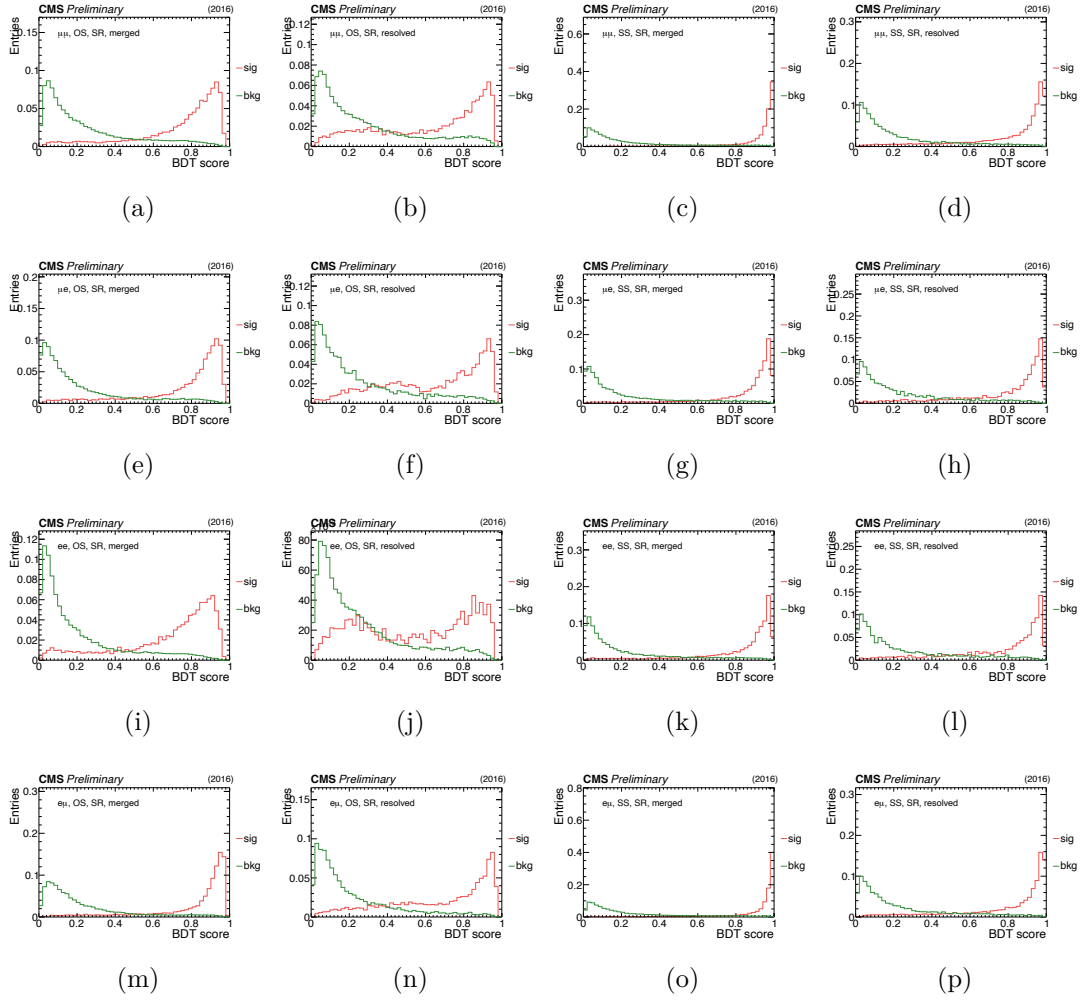


Figure F.14: BDT score distributions of the correlated model, showing good separation between background and signal processes (2016 MC simulation).

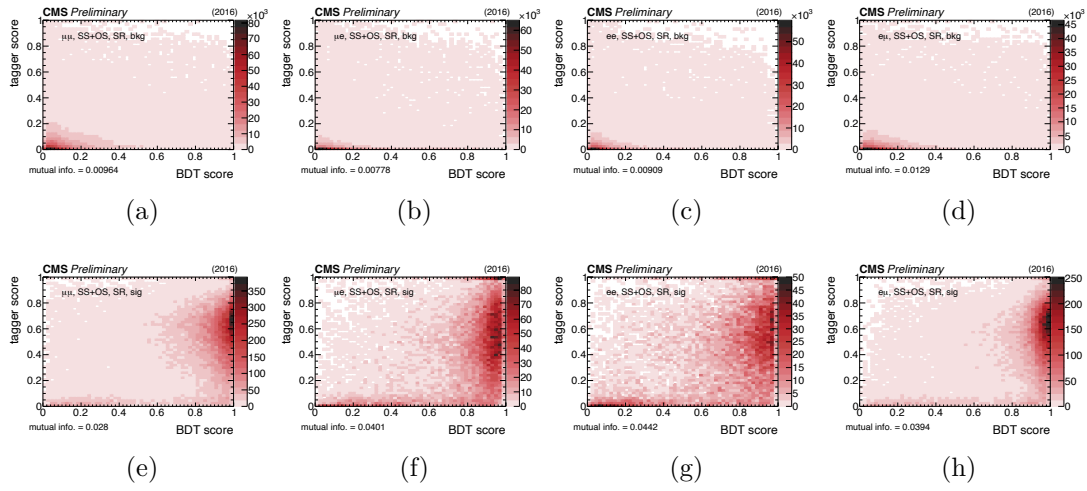


Figure F.15: 2D heatmaps between the BDT score of the correlated model and displaced jet tagger information, categorized by dilepton flavour and combined dilepton charge. Stronger correlation between the tagger information and the BDT output is observed for both background (top row) and signal (bottom row) 2016 MC simulation processes.

F.5 Complete nuisance parameter pulls and impacts

Below are presented the impacts and pulls of the nuisance parameters from the final fit in the unblinded SR of the HNL search.

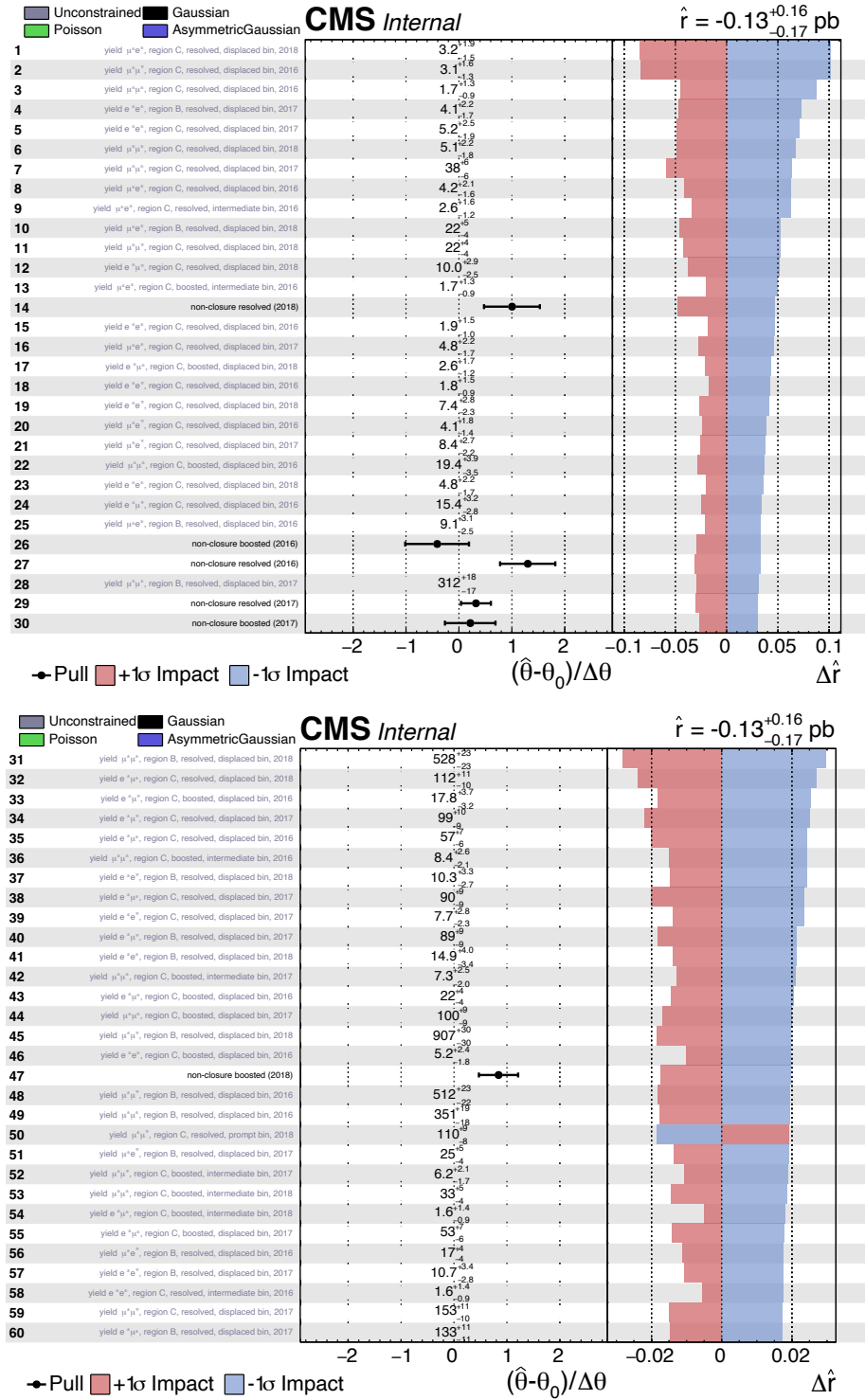


Figure F.16: NP pulls ranked by decreasing impact values. Impacts 1–60 are shown. Signal model used: $m_N = 10$ GeV, $c\tau_0 = 1$ mm, $V_e = V_\mu = V_\tau$.

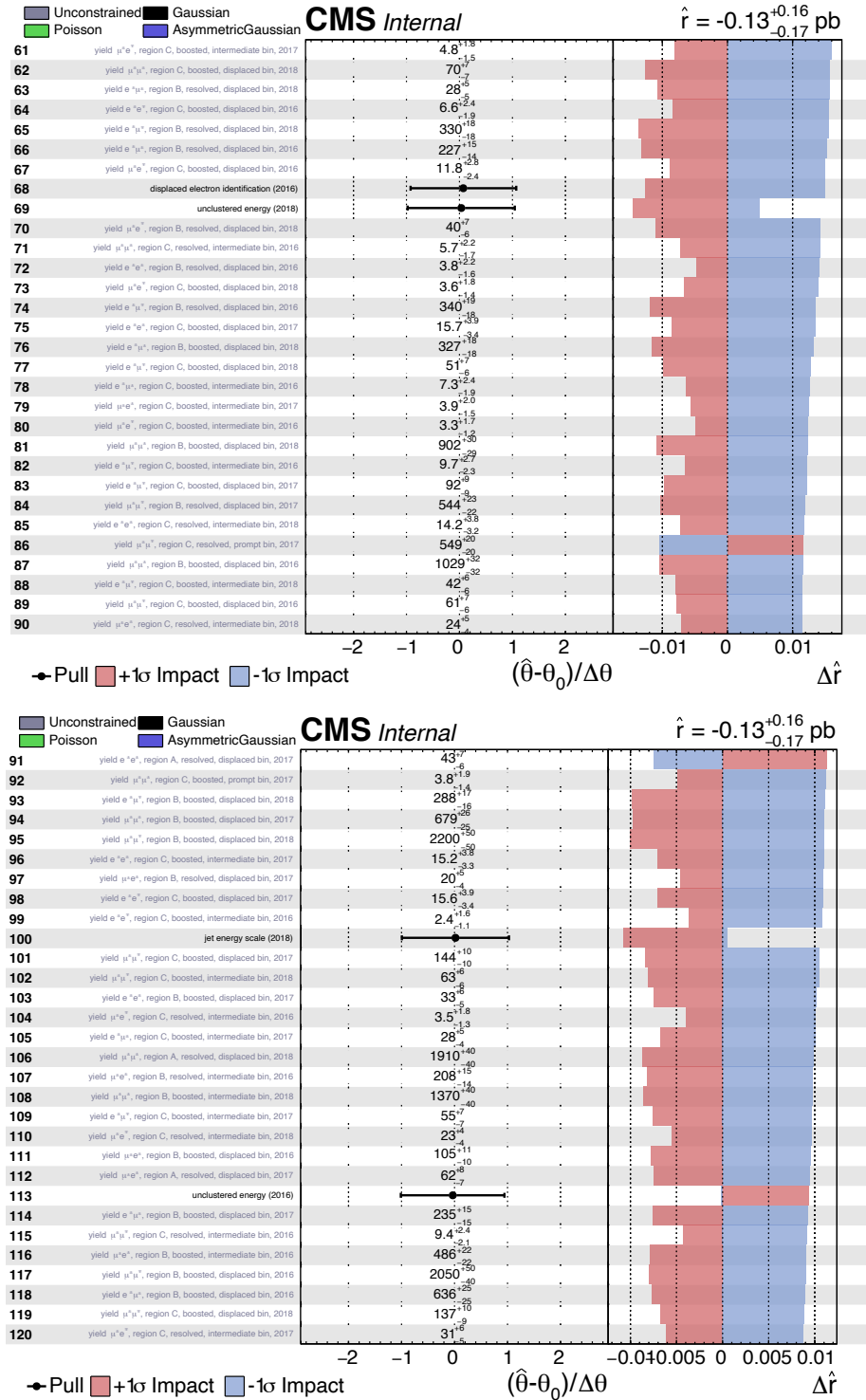


Figure F.17: NP pulls ranked by decreasing impact values. Impacts 61–120 are shown. Signal model used: $m_N = 10$ GeV, $c\tau_0 = 1$ mm, $V_e = V_\mu = V_\tau$.

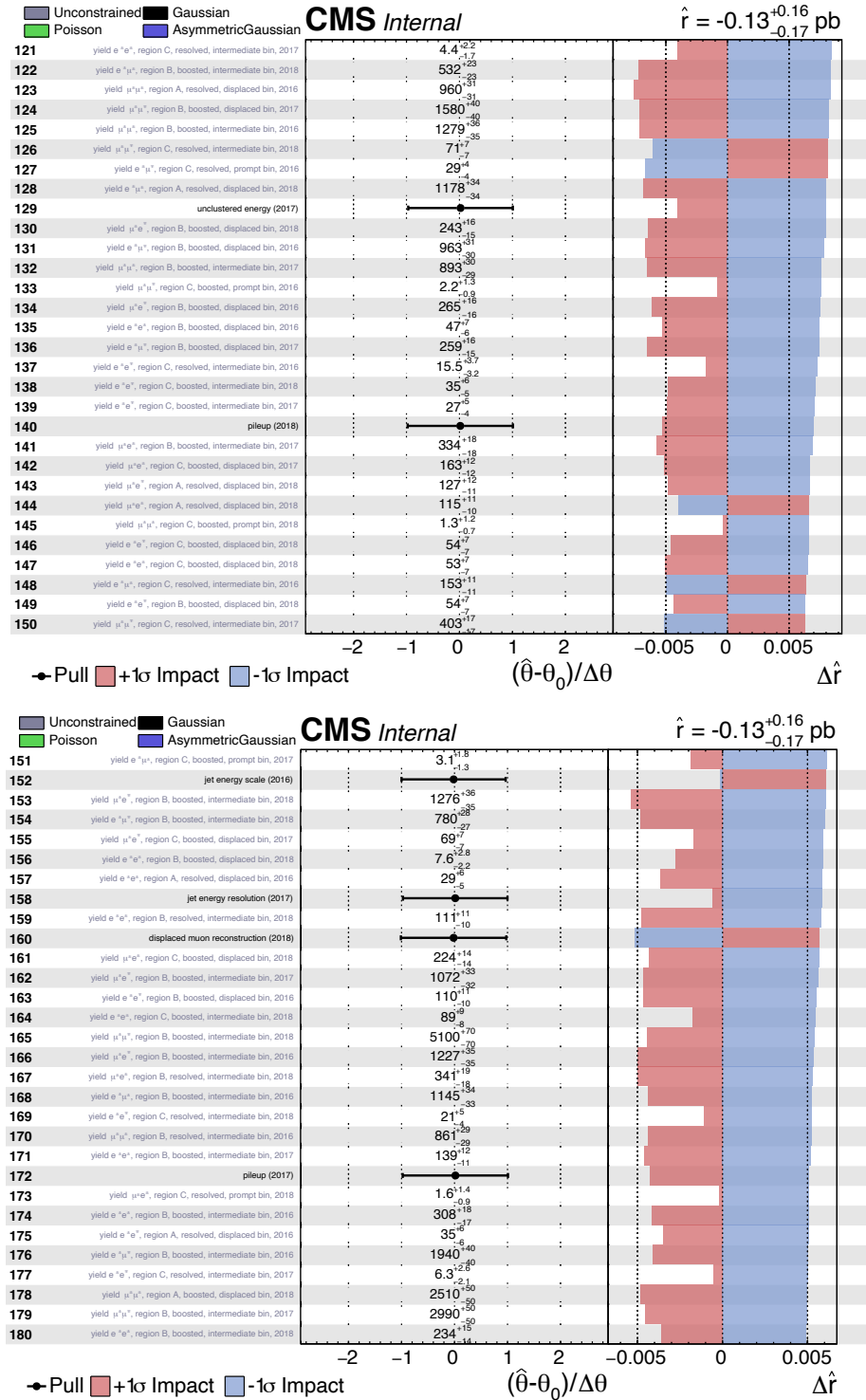


Figure F.18: NP pulls ranked by decreasing impact values. Impacts 121–180 are shown. Signal model used: $m_N = 10$ GeV, $c\tau_0 = 1$ mm, $V_e = V_\mu = V_\tau$.

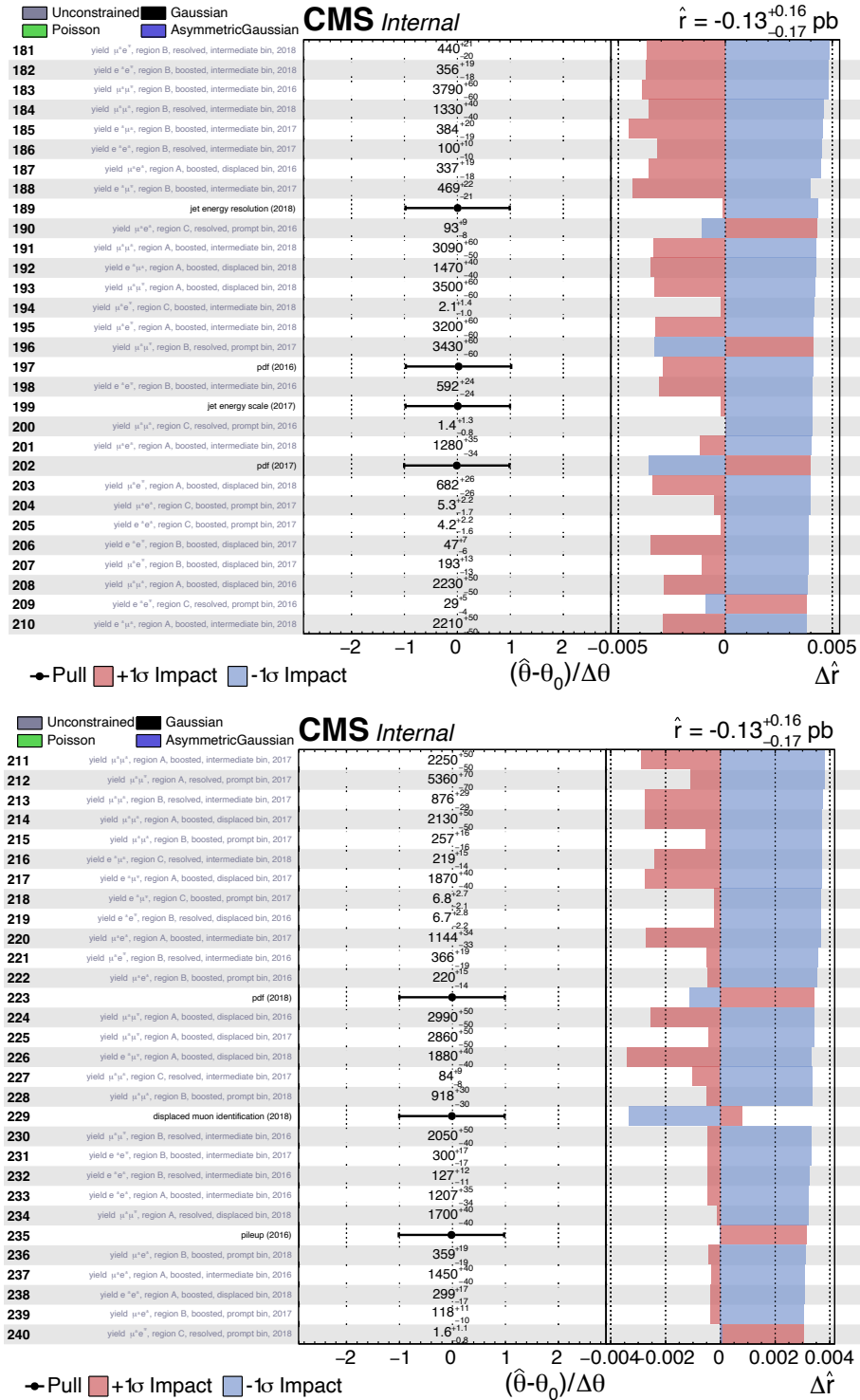


Figure F.19: NP pulls ranked by decreasing impact values. Impacts 181–240 are shown. Signal model used: $m_N = 10$ GeV, $c\tau_0 = 1$ mm, $V_e = V_\mu = V_\tau$.

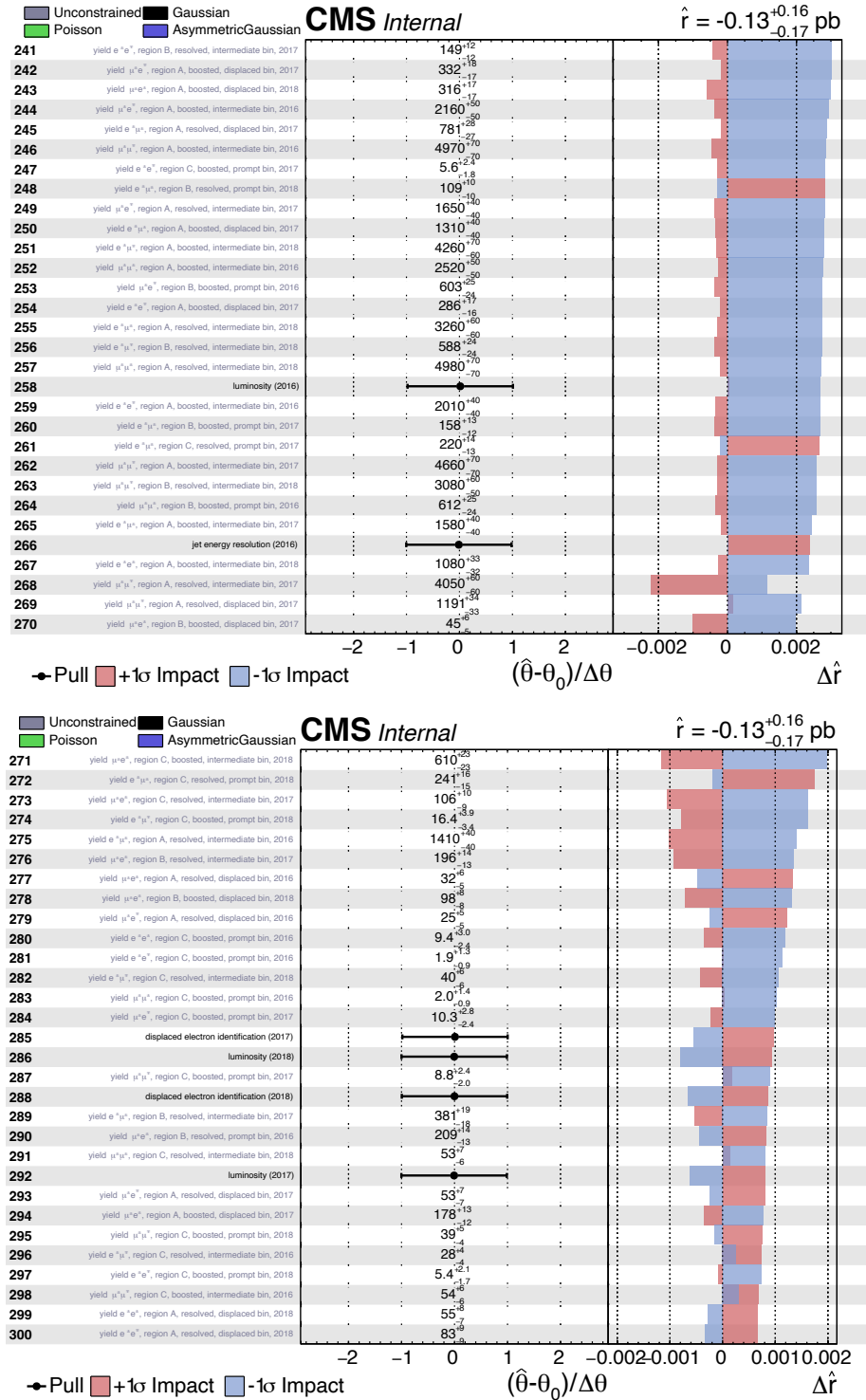


Figure F:20: NP pulls ranked by decreasing impact values. Impacts 241–300 are shown. Signal model used: $m_N = 10$ GeV, $c\tau_0 = 1$ mm, $V_e = V_\mu = V_\tau$.

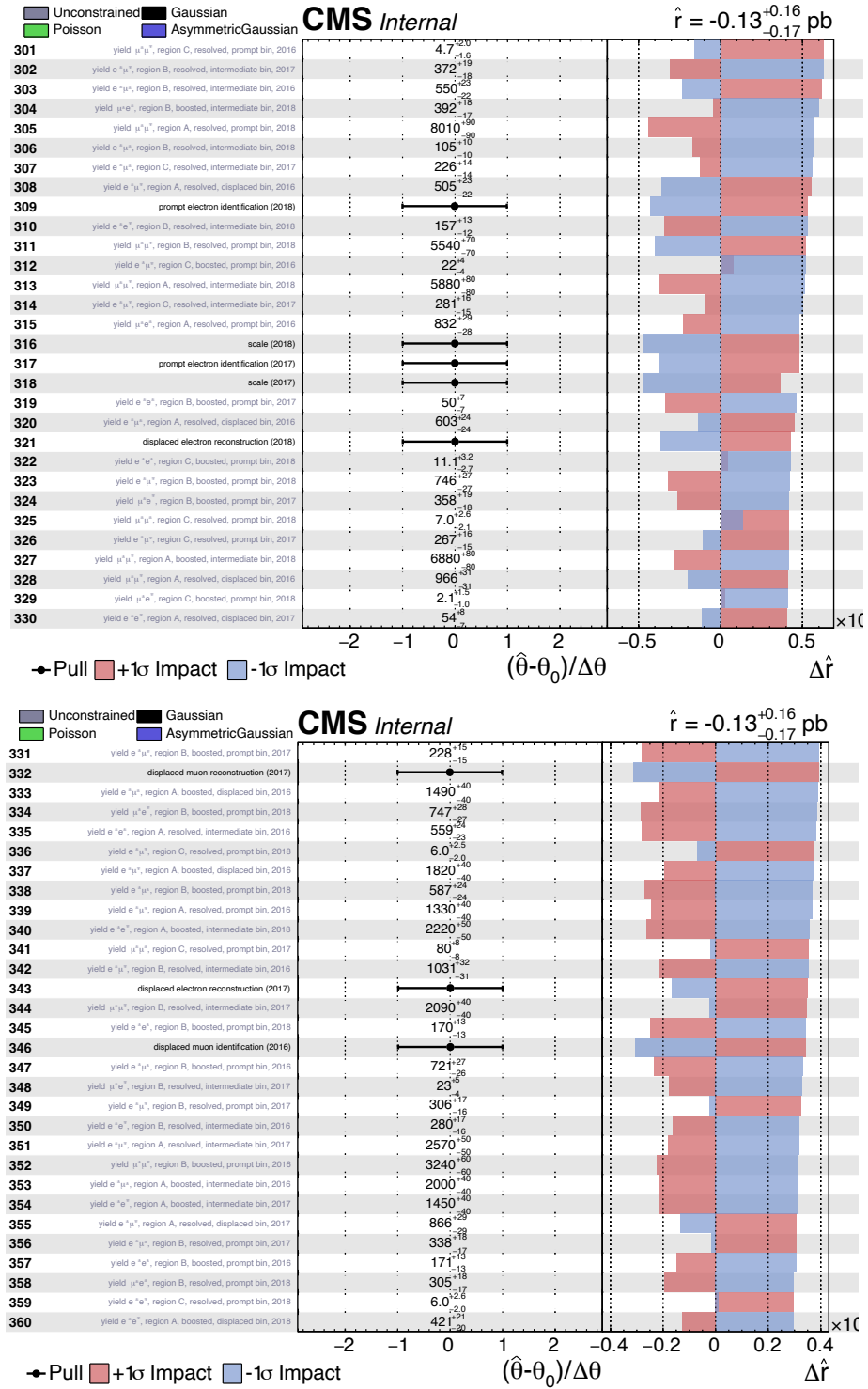


Figure F.21: NP pulls ranked by decreasing impact values. Impacts 301–360 are shown. Signal model used: $m_N = 10$ GeV, $c\tau_0 = 1$ mm, $V_e = V_\mu = V_\tau$.

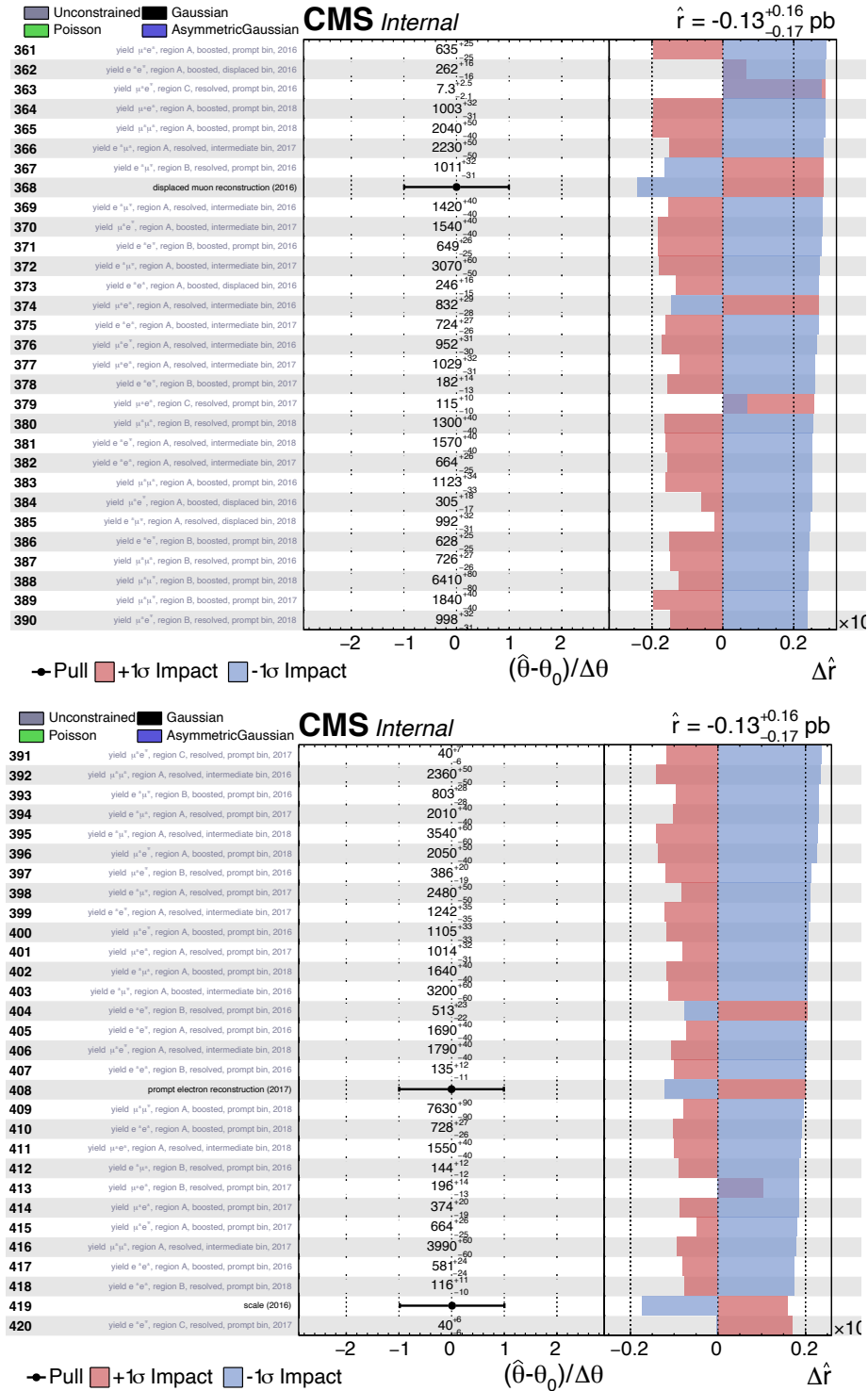


Figure F.22: NP pulls ranked by decreasing impact values. Impacts 361–420 are shown. Signal model used: $m_N = 10$ GeV, $c\tau_0 = 1$ mm, $V_e = V_\mu = V_\tau$.

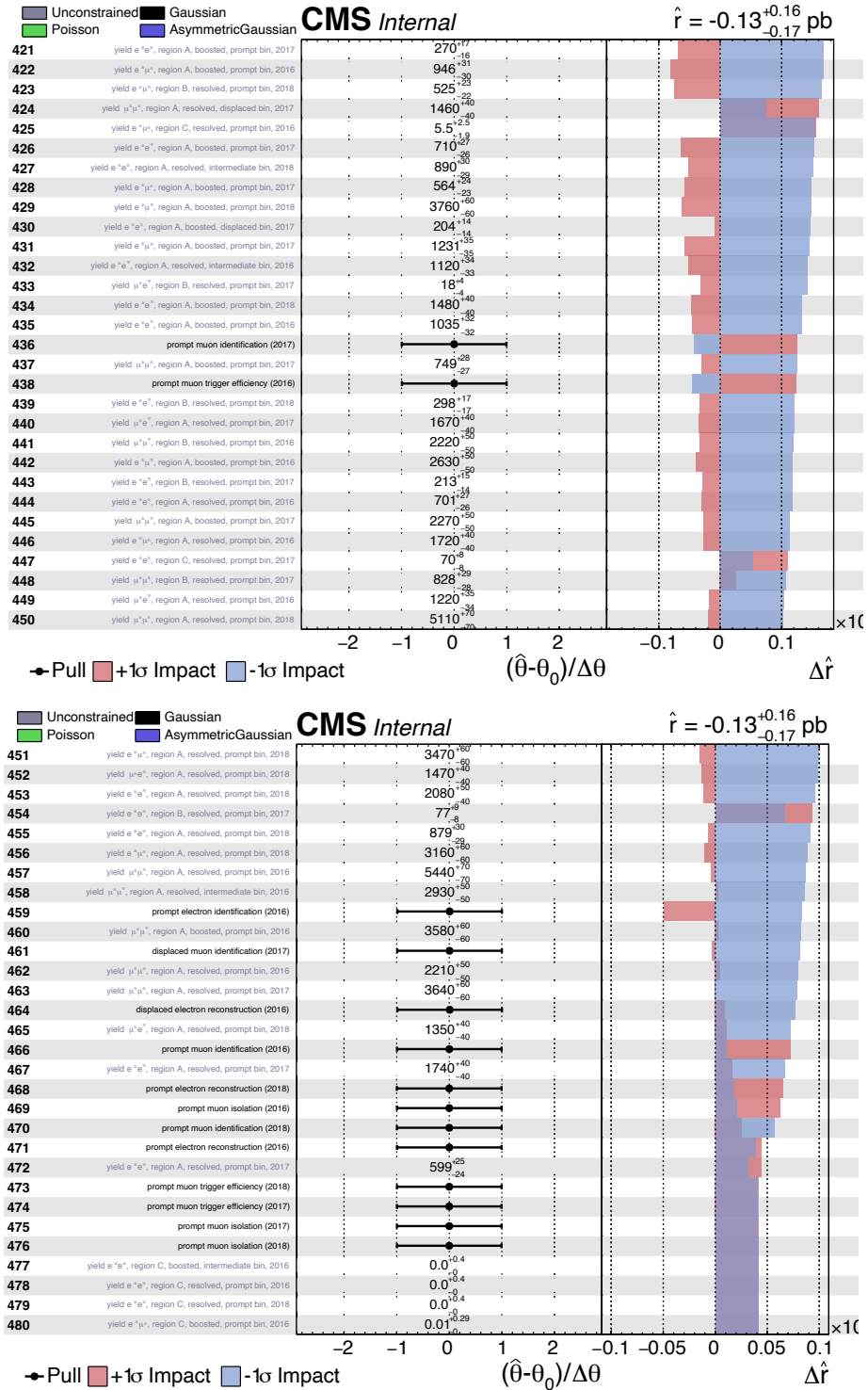


Figure F.23: NP pulls ranked by decreasing impact values. Impacts 421–480 are shown. Signal model used: $m_N = 10 \text{ GeV}$, $c\tau_0 = 1 \text{ mm}$, $V_e = V_\mu = V_\tau$.