

IMPERIAL

IMPERIAL COLLEGE LONDON
DEPARTMENT OF PHYSICS

Search for Hidden Valley dark showers with displaced muons with the CMS experiment

Kai Hong Law

July 14, 2025

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy
in Physics of Imperial College London.

Declaration of Originality

I, Kai Hong Law, declare that the work in this thesis is my own. The work of others has been appropriately referenced. A full list of references is given in the bibliography.

In Chapter 6, I implemented the End-to-End Neural Network Vertexing, and studied its performance and the impact on downstream objects. Chapters 7 and 8 present results of the dark shower analysis, for which I was responsible for all parts except the Monte Carlo generation. The results in Chapter 8 have been approved by the CMS collaboration as preliminary and a journal publication is in progress.

Copyright

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

A search for signatures of a dark analog to quantum chromodynamics is performed. The analysis targets long-lived dark mesons that decay into Standard Model particles with a high branching fraction to muons. It is the first search at the Large Hadron Collider (LHC) that targets the decay of Hidden Valley dark showers into muons. A unique dataset with 10^{10} B meson events is used. It was collected by the CMS experiment at the CERN LHC in 2018 using displaced muon triggers, which have high efficiency for the signal models. Resonant dimuon signatures are searched for, with both pointing and non-pointing topologies. No significant excess is observed beyond the Standard Model expectation. Upper limits on the branching ratio of the Higgs boson decays to dark partons are determined to be as low as about 10^{-4} , at 95% confidence level, surpassing and extending existing limits for the mean proper lifetime of less than approximately 0.1 m and for a mass as low as 2 GeV. First limits are set for extended dark shower models, probing the low-mass region down to 0.33 GeV.

Acknowledgements

I would like to thank my supervisor, Prof. Alexander Tapper, for his continuous guidance throughout my PhD. Thank you for always responding to my questions promptly, and for giving me the opportunities to attend different workshops and conferences. I would also like to thank Matthew Citron, Mikael Mieskolainen, Jaime León Holgado, Benjamin Radburn-Smith, Robert Bainbridge and Chris Brown for the helpful advice and discussions.

Thanks also to Klitos Savva, Jay Odedra, Edward Curtis, Charlotte Knight and Prijith Pradeep for the useful chats.

I am very grateful for the support from the Croucher Foundation through the Croucher Scholarship throughout my PhD. I would also like to thank the Imperial College HEP group for giving me the opportunity to spend a year at CERN.

Contents

1	Introduction	24
2	The Standard Model of particle physics	26
2.1	Introduction	26
2.2	Overview of the Standard Model	26
2.3	Introduction to gauge theories	28
2.4	Quantum electrodynamics	29
2.5	Quantum chromodynamics	30
2.6	Electroweak sector	31
2.7	The Higgs mechanism	32
3	Dark matter and Hidden Valley models	35
3.1	Introduction	35
3.2	Hidden Valley models	35
3.2.1	Vector portal benchmark model	36
3.2.2	Extended benchmark models	39
4	The LHC and the CMS experiment	42
4.1	Introduction	42
4.2	The Large Hadron Collider	42
4.3	LHC performance	44
4.4	The CMS experiment	45
4.4.1	Tracker	46
4.4.2	Electromagnetic Calorimeter	48
4.4.3	Hadron Calorimeter	49
4.4.4	Muon chambers	50
4.4.5	Data acquisition and trigger system	51
5	Event reconstruction	54
5.1	Introduction	54

5.2	Particle Flow reconstruction	54
5.2.1	Track reconstruction	54
5.2.2	Vertex reconstruction	55
5.2.3	Cluster reconstruction	56
5.2.4	Muon reconstruction	57
5.2.5	Electron reconstruction	59
5.2.6	Jet reconstruction	59
5.2.7	Missing energy reconstruction	61
6	Neural Network Vertexing in the Phase-2 upgrade of the Level-1 Trigger	63
6.1	Introduction	63
6.2	Baseline approach for vertex reconstruction	65
6.3	End-to-End Neural Network Vertexing	67
6.3.1	Network architecture and learning	67
6.3.2	Implementation	70
6.3.3	Performance	71
6.3.4	Impact on downstream objects	76
6.3.5	Summary	80
7	Analysis strategy	82
7.1	Introduction	82
7.2	Datasets, triggers and Monte Carlo simulation	82
7.2.1	The B-parking dataset	82
7.2.2	Monte Carlo simulation	85
7.3	Event-level Boosted Decision Tree	89
7.3.1	Introduction	89
7.3.2	Study of jet variables as input variables	92
7.3.3	Training results	94
7.3.4	Study of missing transverse energy input variables	100
7.3.5	Study of spatial relationships of muon secondary vertices as input variables	102
7.3.6	Summary	104
7.4	Event selection	105
7.4.1	Summary	110
7.5	Event categorisation	111
7.5.1	Summary	114
8	Dimuon mass fit, systematic uncertainties, and results	116
8.1	Introduction	116
8.1.1	Fitting technique	117

8.2	Data studies	118
8.2.1	Data and Monte-Carlo comparison	118
8.2.2	Impact of BDT selections on the dimuon mass distribution	121
8.3	Signal parametrisation	125
8.4	Background estimation	130
8.5	Optimisation of the Boosted Decision Tree selection	136
8.6	Systematic uncertainties	138
8.6.1	Overview	138
8.6.2	Systematic uncertainty from the Boosted Decision Tree	139
8.6.3	Systematic uncertainty from displaced muon identification	142
8.6.4	Systematic uncertainty from the triggers	145
8.6.5	Summary	151
8.7	Results and interpretation	153
8.7.1	Signal extraction	153
8.7.2	Branching ratio limits	155
9	Conclusions	160
A	Details of the B-parking triggers	174
B	BDT input variables	176
C	BDT hyperparameters	178
D	Trigger efficiency against l_{xy} for benchmark signal models	179
E	Selection efficiency against $c\tau$ for benchmark signal models	180
F	Impact plots	181
G	Dimuon mass distributions for scenario A and scenario B1	183

List of Figures

3.1	Feynman diagram for the vector portal model. Dark partons $\psi\bar{\psi}$ are first produced from the decay of the SM Higgs boson, which then hadronise to form dark vector mesons $\tilde{\omega}$ and dark scalar mesons $\tilde{\eta}$. The $\tilde{\omega}$ then undergo displaced decay into SM fermions, f	37
3.2	Branching ratio for the decay of the dark meson into different SM fermions for the vector portal as a function of the dark meson mass.	38
3.3	Feynman diagram for the Scenario A model (left) and the Scenario B1 model (right). In these extended models the dark hadronization produces a spectrum of dark mesons, including the dark pions π_1, π_2 and π_3 . The π_3 then decays into SM fermions through the dark photon A' . The A' is a long-lived particle in Scenario A, while the π_3 is a long-lived particle in Scenario B1, as indicated by the green lines.	40
3.4	The multiplicity of dark vector mesons $\tilde{\omega}$ for representative vector portal models (left), and the multiplicity of dark mesons π_3 for representative scenario A and B1 models (right). The fraction of generated events is shown against the multiplicity of dark mesons.	41
4.1	A schematic depiction of the CERN accelerator complex. Image credit to CERN.	43
4.2	The proton-proton luminosity of the LHC, with projections for the HL-LHC up until 2041. Image credit to CERN.	45
4.3	A schematic of the CMS detector.	46
4.4	Graphical representation of the coordinate system used for the CMS experiment.	47

4.5	A schematic cross-section of the CMS tracker in the r - z plane. The tracker is symmetric about $r = 0$ in this view, so only the top half is shown. The centre of the tracker, which is the approximate position of the proton-proton collision point, is indicated by a star. Strip tracker modules which provide 2-D hits are shown by thin, black lines, and those that permit the reconstruction of 3-D hit positions are shown by thick, blue lines. The pixel modules are shown by red lines and they also provide 3-D hits. Within a given layer, each module is shifted slightly in r or z with respect to its neighbouring modules, which allows them to overlap and avoids gaps in the acceptance.	48
4.6	A schematic of the CMS Electromagnetic Calorimeter, showing the barrel, endcap and the preshower subsystems. The barrel section comprises of 36 supermodules, each containing 4 modules.	49
4.7	A schematic view of one quarter of the CMS HCAL, showing its four major components which are the hadron barrel (HB), the hadron endcap (HE), the hadron outer (HO), and the hadron forward (HF) calorimeters. .	50
4.8	A schematic view in the r - z plane of a CMS detector quadrant. The interaction point is located at the lower left corner. The locations of the various muon stations are shown in different colours and labels: drift tubes (DTs) with labels MB, cathode strip chambers (CSCs) with labels ME, resistive place chambers (RPCs) with labels RB and RE, and gas electron multipliers (GEMs) with labels GE. The M denotes muon, B denotes barrel, and E denotes endcap. The magnet yoke is represented by the dark grey areas.	51
5.1	Track reconstruction efficiency as a function of simulated track p_T (left) and production vertex radius (right). The efficiencies are shown successively for each of the twelve tracking iterations, which are indicated by different colours.	56
5.2	Muon reconstruction and identification efficiency in 2015 data and simulation for loose muons with $p_T > 20$ GeV.	58
5.3	Electron identification efficiency measured in data (upper panels) and data-to-simulation efficiency ratio (lower panels) as a function of the electron E_T for a cut-based identification working point (left) and a BDT-based working point (right).	60
5.4	Distributions of p_T^{miss} in $Z \rightarrow \mu^+ \mu^-$ (left) and $Z \rightarrow e^+ e^-$ (right) in data and simulation. The last bin includes all events with $p_T^{\text{miss}} > 195$ GeV. . .	62

6.1	The Phase-2 upgrade design for the CMS Level-1 trigger. The trigger uses information from the calorimeters, the muon spectrometers and the track finder. The correlator trigger makes up of two layers that are dedicated to particle-flow reconstruction. All reconstructed objects are sent to the global trigger, which makes the final Level-1 trigger decision.	64
6.2	Illustration of the upgraded outer tracker p_T module concept. Correlation of signals in closely-spaced sensors allows rejection of low- p_T particles. The selection window shown in green is used to define an accepted stub. .	65
6.3	Vertex reconstruction with the FastHisto algorithm, which uses a p_T weighted histogram of the track z position. The coloured crosses represent the actual position of each track for different types of tracks, which are pileup tracks, fake tracks and primary vertex tracks. The gray markers show the scalar sum of the track p_T in each bin. The black dashed line indicates the vertex position found by the FastHisto algorithm, and the green dashed line shows the position of the true vertex.	66
6.4	Network architecture for the End-to-End Neural Network Vertexing, which learns simultaneously primary vertex regression and track-to-vertex association. Track features are first fed into a dense neural network that learns a weight for each track. The weights are used to fill a histogram in z_0 , which is then passed through a convolutional neural network to perform peak-finding. The z position of the primary vertex is returned with an ArgMax, which finds the maximum bin of the histogram. The track's distance from the primary vertex $\Delta z_{0,i}^{PV}$ and the input track features are then passed to a second dense neural network to learn the probability of the track associated to the regressed primary vertex.	68
6.5	Training loss and validation z_0 mean square error in an example training of the neural network.	71
6.6	z_0 residuals of neural network vertexing trained with the old Kalman Filter tracks for the floating point, quantised (QNN) and quantised pruned (QPNN) neural networks compared to the baseline FastHisto approach. .	72
6.7	Two-dimensional plots showing the z_0 of the reconstructed primary vertex using old Kalman Filter tracks against the z_0 of the true primary vertex for the baseline FastHisto algorithm on the left, and for the quantised neural network on the right. The diagonal strip shows vertices that are correctly reconstructed, while the distribution around the diagonal strip shows vertices which are misreconstructed. The smaller spread of vertices around the diagonal shows a better performance in vertex reconstruction for the neural network compared to the baseline.	73

6.8	z_0 residuals of neural network vertexing trained with the new Kalman Filter tracks for the floating point, quantised (QNN) and quantised pruned (QPNN) neural networks compared to the baseline FastHisto approach. .	74
6.9	Two-dimensional plots showing the z_0 of the reconstructed primary vertex using new Kalman Filter tracks against the z_0 of the true primary vertex for the baseline FastHisto algorithm on the left, and for the quantised neural network on the right. The diagonal strip shows vertices that are correctly reconstructed, while the distribution around the diagonal strip shows vertices which are misreconstructed. The smaller spread of vertices around the diagonal shows a better performance in vertex reconstruction for the neural network compared to the baseline.	74
6.10	Purity against efficiency (left) and receiver operating characteristic (ROC) curves (right) for the baseline cut-based approach and the neural network track-to-vertex association with the old Kalman Filter tracks. The baseline is shown in as a red dot as it represents a single cut, as opposed to varying thresholds for the neural networks. In the ROC plot, the ROC AUC is the area under curve for the neural network approaches. For the baseline, the area is taken as the area of a triangle formed by the origin, the red dot and (1, 1) in the false positive rate against true positive rate plane.	75
6.11	Purity against efficiency (left) and receiver operating characteristic (ROC) curves (right) for the baseline cut-based approach and the neural network track-to-vertex association with the new Kalman Filter tracks. The baseline is shown in as a red dot as it represents a single cut, as opposed to varying thresholds for the neural networks. In the ROC plot, the ROC AUC is the area under curve for the neural network approaches. For the baseline, the area is taken as the area of a triangle formed by the origin, the red dot and (1, 1) in the false positive rate against true positive rate plane.	76
6.12	Trigger efficiency of PUPPI MET as a function of generated MET at a fixed rate of 18 kHz using the vertex from neural network vertexing, the baseline FastHisto algorithm and Monte-Carlo truth-matched tracks respectively. In the case of truth tracks, the truth-matched tracks are input to the FastHisto algorithm to obtain the vertex, and only the truth tracks are associated to the vertex. The plot on the right shows a direct comparison between the efficiencies from neural network vertexing and the FastHisto algorithm.	77

6.13	Trigger rate of PUPPI MET as a function of offline p_T threshold, making comparisons between the neural network and the FastHisto algorithm (left), and between the FastHisto and truth tracks (right). In the case of truth tracks, Monte Carlo truth-matched tracks are input to the FastHisto algorithm to obtain the vertex, and only the truth tracks are associated to the vertex. The fluctuations at lower rates are due to statistical errors. . . .	78
6.14	Matching efficiency of generated tau with the Level-1 neural network PUPPI tau in the barrel (left) and the endcap (right) using the FastHisto algorithm (blue) and the neural network vertexing (orange) respectively. A threshold of 0.1 is used for the association network in the neural network vertexing. A clear improvement in the matching efficiency is observed for the neural network vertexing.	79
6.15	Trigger efficiency of the neural network PUPPI tau in the barrel (left) and the endcap (right) respectively. A threshold of 0.1 is used for the association network in the neural network vertexing. A Level-1 p_T threshold of 30 GeV is applied on the taus for the FastHisto algorithm, while the p_T threshold is tuned for the neural network vertexing to obtain roughly the same efficiency as the FastHisto.	80
6.16	Trigger rate of the neural network PUPPI tau as a function of online threshold in the barrel (left) and the endcap (right) using the FastHisto algorithm and the neural network vertexing respectively. A threshold of 0.1 is used for the association network in the neural network vertexing. . .	80
7.1	The typical Run 2 data flow during 2018, which shows the different data acquisition strategies, including the parking data stream. A value of $\mathcal{L}_{\text{inst}} = 1.2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ over a typical 2018 fill is used, which corresponds to an average pileup of 38.	83
7.2	The pileup distribution for the 2018 B-parking dataset. Contributions from each individual trigger combination are shown. The areas of the histograms are normalised to the number of events recorded by each trigger.	85
7.3	The invariant mass distribution of oppositely charged dimuons originating from a common vertex, obtained using a subset of the 2018 B-parking dataset. The contributions from each trigger combination are shown. Different Standard Model resonances are labelled.	86
7.4	Distributions of some variables that are used in the BDT training for the QCD background and benchmark signal models.	90

7.5	The receiver operating characteristic (ROC) curves of a range of machine learning techniques. The plot compares the performance of the different machine learning models to that of cut-based selections (which require at least one secondary vertex, at least one jet and at least four muons). The XGB (XGBoost) model is found to give the best performance, as indicated by the largest area under its ROC curve.	92
7.6	The loss (left) and the ROC AUC (right) against the epochs for the vector portal training. They are shown for both the training and the validation datasets.	93
7.7	ROC curves showing the performances of the BDT trained with variables from muons, secondary vertices and muon secondary vertices (XGB-NOJETS) compared to the BDT trained with additional jet variables (XGB). The BDTs are trained with the vector portal, scenario A and scenario B1 signal model respectively.	94
7.8	BDT output distributions for the signal and the background from the training performed with the vector portal, scenario A and scenario B1 model respectively. The BDTs use variables from muons, secondary vertices and muon secondary vertices. $C = 0$ denotes the distribution for the QCD background sample, while $C = 1$ denotes the distribution for the signal sample.	95
7.9	Feature importance plot for the vector portal training showing the top 50 features that have the highest "gain" scores, which are features that bring the most improvement to the performance of the BDT when it makes selections on those features. The numbers in round brackets show the feature importance scores for each feature. The x -axis is truncated for readability. "muonSV" refers to muon secondary vertex. The number at the end of the feature name corresponds to the index of the object.	97
7.10	Distributions of example features which provide the best discriminating power between the signal and the background. The distributions are shown for the vector portal signal, the QCD background and a small subset of data. The signal and the data are reweighted to give the same number of weighted events as the QCD background. The plots are normalised to the same number of events, but not necessarily the same number of objects.	98

7.11	Distributions of example features which provide the best discriminating power between the signal and the background. The distributions are shown for the vector portal signal, the QCD background and a small subset of data. The signal and the data are reweighted to give the same number of weighted events as the QCD background. The plots are normalised to the same number of events, but not necessarily the same number of objects.	99
7.12	Data versus Monte Carlo distributions of example features which provide great discriminating power between the signal and the background.	100
7.13	Distributions of MET variables. The distributions are shown for the vector portal signal, the QCD background and a small subset of data. The signal and the data are reweighted to give the same number of weighted events as the QCD background.	101
7.14	ROC curves showing the performances of the BDTs trained with or without variables of missing transverse energy. The trainings of both BDTs are performed with the vector portal, scenario A and scenario B1 signal models respectively.	102
7.15	Distributions of ΔR between muon secondary vertices. The plots show the four smallest ΔR between pairs of muon secondary vertices in each event. The distributions are shown for the vector portal signal, the QCD background and a small subset of data. The signal and the data are reweighted to give the same number of weighted events as the QCD background.	103
7.16	ROC curves showing the performances of the BDTs trained with or without input variables from the ΔR between pairs of muon secondary vertices. The trainings of both BDTs are performed with with the vector portal, scenario A and scenario B1 signal models respectively.	105
7.17	Distribution of BDT score (left) and BDT selection efficiency (right) for the QCD background and selected signal models.	107
7.18	Distribution of $\Delta R(\mu, \mu)$ for the QCD background and some benchmark signal models. All events passing the baseline selections are used.	108
7.19	Distributions of the positions of muon secondary vertices in the $x - y$ plane for a subset of data events which pass a loosened BDT selection, before (top) and after (bottom) the material veto is applied.	109
7.20	Distributions of muon secondary vertex transverse displacement (top) and z position (bottom) for a subset of data. The shaded bands in the top plot indicate the regions in transverse displacement (l_{xy}) where material veto is applied, and the bottom plot shows the selections applied to the z position of the muon secondary vertex after the l_{xy} selections are applied.	110

7.21	Categorising events into single vertex events (left) and multi vertex events (right) for events that pass all selections. The distribution for the QCD background uses the shape of the distribution from a looser BDT selection. For multi vertex events, the SV with the lowest χ^2 is chosen among the pairs of SVs of mass within 3% of each other.	112
7.22	Distribution of muon SV transverse displacement for the QCD background and benchmark signal models (top), and for selected signal models of different lifetimes (bottom). All events passing the baseline selections are displayed.	113
7.23	Illustration of the pointing angle of the muon SV, which is defined as the angle between the momentum vector of the dimuon system and the displacement vector from the primary vertex to the secondary vertex. . . .	114
7.24	Distribution of muon SV pointing angle for the QCD background and benchmark signal models. All events passing the baseline selections are displayed.	114
8.1	Distributions of the BDT output score (left) and the dimuon invariant mass (right) for a subset of data and the QCD background Monte Carlo. The BDT used in the left plot is trained for the vector portal signal model. Events shown pass all selections except the BDT selection and the material veto. The "base category" label means all of the twelve analysis categories combined.	118
8.2	Dimuon mass distributions in the six single vertex categories for a subset of data, the QCD background Monte Carlo and a benchmark signal model from scenario A. Events shown pass all selections except the BDT selection and the material veto.	119
8.3	Dimuon mass distributions in the six multi vertex categories for a subset of data, the QCD background Monte Carlo and a benchmark signal model from scenario A. Events shown pass all selections except the BDT selection and the material veto.	120
8.4	Distributions of some variables of the muon secondary vertex with the lowest χ^2 for a subset of data and the QCD background Monte Carlo. Events shown pass all selections except the BDT selection and the material veto. These plots are normalised to unity to compare the shapes of the data and MC distributions.	121
8.5	Fits of the ratio of the dimuon mass distribution with different BDT selections to the inclusive distribution for different signal mass hypotheses using a subset of data. The χ^2/ndf for each fit is also shown. No sculpting of the mass distribution is observed.	122

8.6	The χ^2/ndf of the fits to the ratio of the dimuon mass distribution in a subset of data with different BDT selections to that without the selections are plotted for different signal mass hypotheses.	123
8.7	Fits of the ratio of same-sign dimuon mass distribution with the analysis BDT selection to the inclusive distribution for different signal mass hypotheses, using a subset of data. The χ^2/ndf for each fit is also shown. No sculpting of the dimuon mass is observed.	124
8.8	Fits of the ratio of same-sign dimuon mass distribution with the analysis BDT selection to the inclusive distribution for different signal mass hypotheses, using a subset of data. The χ^2/ndf for each fit is also shown. No sculpting of the dimuon mass is observed.	125
8.9	Fits of signal dimuon mass distributions for events in single vertex category 4 ($1 < l_{xy} < 10$ cm, pointing angle > 0.2). Events shown are for the vector portal models with $m_{\tilde{\omega}} = 5$ GeV and different lifetimes which pass all selections. The χ^2/ndf and the HWHM for each fit are also shown. . .	126
8.10	Fits of signal dimuon mass distributions for events in single vertex category 3 ($1 \text{ cm} < l_{xy} < 10$ cm, pointing angle < 0.2). Events shown are for the scenario A models with $m_{\pi 3} = 4$ GeV, $m_{A'} = 1.33$ GeV and different lifetimes which pass all selections. The χ^2/ndf and the HWHM for each fit are also shown.	127
8.11	Fits of signal dimuon mass distributions for events in single vertex category 4 ($1 \text{ cm} < l_{xy} < 10$ cm, pointing angle > 0.2). Events shown are for the scenario A models with $m_{\pi 3} = 4$ GeV, $m_{A'} = 1.33$ GeV and different lifetimes which pass all selections. The χ^2/ndf and the HWHM for each fit are also shown.	127
8.12	Relative resonance width σ as a function of the dimuon mass for representative lifetime models in the vector portal (upper left), scenario A (upper right) and scenario B1 (lower).	128
8.13	Relative resonance width σ as a function of the dimuon mass in different event categories for representative lifetime models in the vector portal. . .	129
8.14	Background fits of dimuon mass distributions in data at example signal mass hypotheses in the single vertex category 3.	130
8.15	Bias values for $r_{\text{injected}} = 5$ against the dimuon mass for three representative lifetime hypotheses in scenario A.	132
8.16	Goodness-of-fit test (using the saturated model) p-values as a function of mass in the single vertex categories. Each point is coloured red, orange, blue or green if the number of observed events in the mass window is 0, 1-9, 10-99 or above 99 respectively.	132

8.17	Goodness-of-fit test (using the saturated model) p-values as a function of mass in the multi vertex categories. Each point is coloured red, orange, blue or green if the number of observed events in the mass window is 0, 1-9, 10-99 or above 99 respectively.	133
8.18	Dimuon mass distributions in the single vertex categories for data and two benchmark signal models in the vector portal, using events that pass all selections. The shaded regions indicate mass regions of known Standard Model resonances, which are masked in the search. The masked regions are detailed in Table 8.1.	134
8.19	Dimuon mass distributions in the multi vertex categories for data and two benchmark signal models in the vector portal, using events that pass all selections. The shaded regions indicate mass regions of known Standard Model resonances, which are masked in the search. The masked regions are detailed in Table 8.1.	135
8.20	95% confidence level expected upper limits against the BDT background efficiency for $m = 2, 5, 15$ and 20 GeV of the vector portal model. Representative lifetime models are studied for each mass hypothesis.	137
8.21	95% confidence level expected upper limits against the BDT background efficiency for $m = 0.33, 0.4, 0.67$ and 1.33 GeV of the scenario A model. Representative lifetime models are studied for each mass hypothesis. . . .	137
8.22	95% confidence level expected upper limits against the BDT background efficiency for $m = 0.33, 0.4, 0.67$ and 1.33 GeV of the scenario B1 model. Representative lifetime models are studied for each mass hypothesis . . .	138
8.23	BDT score distributions in the J/ψ mass region for QCD Monte Carlo and a subset of data using the BDT trained with the vector portal signal model.	140
8.24	Fit for the passing events (left) and the failing events (right) for the QCD Monte Carlo.	140
8.25	Fit for the passing events (left) and the failing events (right) for a subset of data.	140
8.26	BDT efficiencies and scale factors as a function of the background efficiency for the vector portal BDT.	141
8.27	BDT efficiencies and scale factors as a function of the background efficiency for the scenario A BDT.	141
8.28	BDT efficiencies and scale factors as a function of the background efficiency for the scenario B1 BDT.	142
8.29	Fit for the passing probes (left) and the failing probes (right) for Monte Carlo in the bin of $0.001 \text{ cm} < d_{xy} < 0.1 \text{ cm}$ and $4 \text{ GeV} < p_T < 6 \text{ GeV}$.	143
8.30	Muon identification efficiency for data in bins of transverse momentum and transverse displacement. A subset of B-parking data is used.	144

8.31	Muon identification efficiency for Monte Carlo in bins of transverse momentum and transverse displacement. A Monte Carlo sample of $B \rightarrow J/\psi K$ is used.	144
8.32	Muon identification scale factor in bins of transverse momentum and transverse displacement.	145
8.33	Fit for the passing probes (left) and the failing probes (right) for data in the bin of $0.3 \text{ cm} < d_{xy} < 0.4 \text{ cm}$ and $10 \text{ GeV} < p_T < 16 \text{ GeV}$	146
8.34	Trigger efficiency of the HLT_Mu9_IP6 trigger for data in bins of transverse momentum and transverse impact parameter.	147
8.35	Trigger efficiency of the HLT_Mu9_IP6 trigger for Monte Carlo in bins of transverse momentum and transverse impact parameter.	147
8.36	Trigger scale factors for the HLT_Mu9_IP6 trigger in bins of transverse momentum and transverse impact parameter.	148
8.37	Trigger efficiencies and scale factors for the HLT_Mu9_IP6 trigger in bins of transverse impact parameter.	148
8.38	Trigger efficiency of the HLT_Mu9_IP6 trigger for data in bins of muon transverse momentum and transverse displacement of the muon secondary vertex.	149
8.39	Trigger efficiency of the HLT_Mu9_IP6 trigger for Monte Carlo in bins of muon transverse momentum and transverse displacement of the muon secondary vertex.	149
8.40	Trigger scale factors for the HLT_Mu9_IP6 trigger in bins of muon transverse momentum and transverse displacement of the muon secondary vertex.	150
8.41	Trigger efficiencies and scale factors for the HLT_Mu9_IP6 trigger in bins of transverse displacement of the muon secondary vertex.	150
8.42	Trigger scale factors in bins of transverse momentum and transverse impact parameter significance of the muon for all B-parking triggers.	151
8.43	Background fit in a mass window centred at 0.67 GeV (top) for scenario A and a mass window centred at 1.33 GeV (bottom) for scenario B1, both in the single vertex category with $1 \text{ cm} < l_{xy} < 10 \text{ cm}$ and pointing angle < 0.2 . The signals expected for a representative scenario A model (top) and a scenario B1 model (bottom) are shown together with the background fits. The signals are normalised to a branching ratio of 0.01.	153
8.44	Pre-fit and post-fit plots for example categories for scenario A, $m_\pi = 1.0 \text{ GeV}$, $m_A = 0.33 \text{ GeV}$, $c\tau = 10 \text{ mm}$. The signal is shown in green, the background is shown in red, and signal+background is shown in purple. The signal is normalised to a branching ratio of 0.01 in the pre-fit.	154

8.45	Pre-fit and post-fit plots for example categories for scenario B1, $m_{\pi_3} = 4.0$ GeV, $m_A = 1.33$ GeV, $c\tau = 10$ mm. The signal is shown in green, the background is shown in red, and signal+background is shown in purple. The signal is normalised to a branching ratio of 0.01 in the pre-fit.	155
8.46	Exclusion limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ for representative mass hypotheses for the vector portal model. It is assumed that $m_{\tilde{\omega}} = \tilde{\Lambda} = m_{\tilde{\eta}}$, where $m_{\tilde{\omega}}$, $\tilde{\Lambda}$ and $m_{\tilde{\eta}}$ are the mass of the dark sector spin-one meson, the dark sector confinement scale, and the mass of the dark sector spin-zero meson respectively.	156
8.47	Exclusion limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ for representative mass hypotheses for the scenario A model. It is assumed that $m_{\eta} = \tilde{\Lambda} = 4m_{\pi_2}$ and $\sin \theta = 0.1$, where m_{η} is the mass of the dark sector pseudoscalar and θ is the mixing angle parametrising the isospin violation. The branching ratio $\mathcal{B}(\pi_3 \rightarrow A'A')$ is assumed to be one.	157
8.48	Exclusion limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ for representative mass hypotheses for the scenario B1 model. It is assumed that $m_{\eta} = \tilde{\Lambda} = 4m_{\pi_2}$ and $\sin \theta = 0.1$, where m_{η} is the mass of the dark sector pseudoscalar and θ is the mixing angle parametrising the isospin violation. The branching ratio $\mathcal{B}(\pi_3 \rightarrow A'A')$ is assumed to be one.	158
8.49	Observed limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ for the vector portal model from this analysis compared to those from Ref. This analysis targets decays into dimuons using 41.6 /fb of proton-proton collision data, while Ref. is an inclusive search with muon detector showers using 138 /fb of proton-proton data that is sensitive to all decay modes except for muons.	159
D.1	Trigger efficiency for the HLT_Mu9_IP6 trigger against the transverse displacement l_{xy} of the generated vertex in different signal models. Both muons in the vertex are required to have $p_T > 9$ GeV and $ \eta < 1.5$. The efficiency drops quickly after about 15 cm.	179
E.1	Selection efficiency against the mean proper lifetime for benchmark signal models.	180
F.1	Impact plot made using toys, without signal injection. The plot shows the fits of the nuisance parameters, and the impacts of the nuisance parameters on the signal strength. The scenario A model point with $m_{\pi_3} = 4$ GeV, $m_{A'} = 0.4$ GeV, $c\tau = 60$ mm is used for the study.	181

F.2	Impact plot made using toys, with signal injection. The plot shows the fits of the nuisance parameters, and the impacts of the nuisance parameters on the signal strength. The scenario A model point with $m_{\pi 3} = 4$ GeV, $m_{A'} = 0.4$ GeV, $c\tau = 60$ mm is used for the study.	182
G.1	Dimuon mass distributions in the single vertex categories for data and two benchmark signal models in scenario A. The shaded regions indicate mass regions of known SM resonances, which are masked in the search. .	183
G.2	Dimuon mass distributions in the multi vertex categories for data and two benchmark signal models in scenario A. The shaded regions indicate mass regions of known SM resonances, which are masked in the search.	184
G.3	Dimuon mass distributions in the single vertex categories for data and two benchmark signal models in scenario B1. The shaded regions indicate mass regions of known SM resonances, which are masked in the search. .	185
G.4	Dimuon mass distributions in the multi vertex categories for data and two benchmark signal models in scenario B1. The shaded regions indicate mass regions of known SM resonances, which are masked in the search. .	186

List of Tables

2.1	The twelve fundamental fermions in the Standard Model, divided into three generations of quarks and leptons. The electric charge and the current mass of each particle are shown.	27
2.2	The three fundamental forces described by the Standard Model, and their corresponding gauge bosons.	27
6.1	Resource usage and latency of one SLR of a Xilinx VU13P running at 360 MHz for the quantised pruned (QP) weight network, quantised (Q) pattern network and quantised pruned (QP) association network in the end-to-end neural network vertexing. The resource usage is estimated by a Vivado implementation of the networks, while the latency and initiation interval are estimated from a simulation. Post-training quantisation is applied to reduce the resource usage.	76
7.1	Trigger configurations in the 2018 B-parking dataset, defined by unique combinations of Level-1 (L1) p_T , high-level trigger (HLT) p_T and transverse impact parameter significance (IP_{sig}) thresholds. The integrated luminosity (\mathcal{L}_{int}), the mean number of pileup interactions, and the number of events recorded by each combination are shown as the total over the periods for which each combination provided the lowest enabled L1 p_T threshold.	85
7.2	Model parameters of the different classes of signal models interpreted by the analysis.	86
7.3	Signal models with different masses and lifetimes that are produced in the analysis.	87
7.4	Cross-sections for the p_T -binned QCD samples that are used in the analysis.	87
7.5	The trigger efficiency \times acceptance for representative benchmark models in the vector portal, scenario A and scenario B1.	88

7.6	Training samples for the event-level BDT. A BDT is trained for each class of signal model. The number of events passing pre-selections that are used for the BDT is shown for each signal model. The training/testing/validation split is 60:30:10.	93
7.7	Full event categorisation for the analysis. Events are first divided into the single vertex and multi vertex categories. Each of the categories is then subdivided into 6 categories based on the transverse displacement and pointing angle of the muon SV with the minimum χ^2 , resulting in 12 categories in total.	115
8.1	List of Standard Model resonances and the corresponding mass windows that are "masked" in the analysis.	136
8.2	The BDT background efficiency working points used for the different mass models in the vector portal after optimisation.	138
8.3	Selections used for the tag-and-probe technique to measure the displaced muon identification efficiency.	142
8.4	Selections used for the tag-and-probe technique to measure trigger efficiency.	146
8.5	Summary of the systematic uncertainties applied to the signal.	152
C.1	Hyparameters used for the BDT training	178

Chapter 1

Introduction

One of the biggest open questions in particle physics is the nature of dark matter. Various astrophysical observations provide strong evidence for the existence of a form of matter that does not interact with the electromagnetic force but has mass, known as dark matter. A compelling example is the measurement of galaxy rotation curves, which show that the rotation velocities are higher than expected, strongly indicating the existence of dark matter [1]. About 27% of the energy content of our Universe is believed to be made up of dark matter, compared to only 5% being visible matter [1, 2, 3, 4].

Many theories beyond the Standard Model (SM) of particle physics predict the existence of dark matter which can be probed with particle colliders. The Hidden Valley model is a class of model that gives long-lived dark particles with high multiplicity, and therefore has a rich phenomenology. This thesis presents a search for Hidden Valley dark showers decaying into displaced muons with the Compact Muon Solenoid (CMS) experiment. This is the first search at the Large Hadron Collider (LHC) that targets the decay of dark showers into muons.

In this thesis, an overview of the SM of particle physics is provided in Chapter 2. Dark matter is not explained by the SM, which motivates the need for theories beyond the SM. Chapter 3 gives an introduction to Hidden Valley models, in which the SM is extended by a non-Abelian gauge group. The different benchmark models considered in the analysis are presented.

An overview of the LHC and the different subdetectors of the CMS experiment is given in Chapter 4. The reconstruction of particles in the CMS experiment is discussed in Chapter 5. The LHC will be upgraded to provide an unprecedented level of luminosity and pileup, which necessitate upgrades of the CMS detector to cope with the challenging detector environment. Chapter 6 discusses an upgrade that involves performing vertexing using neural networks in the first level of the trigger system, which is used to make fast

selection of collision events. Results showing the performance of the vertexing, and its impact on the trigger efficiency and rate are presented.

The analysis strategy is discussed in Chapter 7. A unique B-parking dataset is used for the search, which utilises delayed reconstruction to record a large number of displaced decays into muons. Event-level Boosted Decision Trees are trained in the analysis to discriminate between the signal and the background, and they show high capability in rejecting background. In Chapter 8, studies with data are presented, as well as details of the signal parametrisation and background estimation. Different sources of systematic uncertainties are then discussed. Upper limits are imposed on the branching ratio of the Higgs boson decaying into dark partons, which will be presented for different benchmark models considered in the analysis. Finally, conclusions are drawn in Chapter 9.

Chapter 2

The Standard Model of particle physics

2.1 Introduction

In this Chapter, a brief overview of the Standard Model of particle physics is provided. The fundamental particles, and the relationships and interactions between them are discussed. The review is based on Ref. [5].

2.2 Overview of the Standard Model

The Standard Model of particle physics encapsulates the current understanding of elementary particles and the interactions between them, which are forces described by the exchange of particles. It agrees with almost all current experimental data and represents one of the successes of modern physics.

Normal matter is made up of atoms, which are bound states of negatively charged electrons (e^-) which orbit around a central nucleus consisting of positively charged protons (p) and electrically neutral neutrons (n). At higher energy scales, protons and neutrons are found to be bound states of fundamental particles referred to as quarks. The proton consists of two up-quarks and a down-quark, $p(uud)$, while the neutron consists of two down-quarks and an up-quark, $n(duu)$. Another fundamental particle encountered in nature comes from nuclear β -decays of certain isotopes and the nuclear fusion processes in the Sun, and it is known as the electron neutrino (ν_e), which is nearly massless.

The up-quark, the down-quark, the electron and electron neutrino are collectively known as the first generation. Further generations of particles were observed by particle experiments - for each of the four first-generation particles, there are two copies which differ only in their masses, leading to the second and third generations. Therefore the matter content of the Universe appears to be described by twelve fundamental spin-half parti-

cles, which are summarised in Table 2.1. The particles can be divided into quarks and leptons.

The quarks and leptons interact through the four fundamental forces, which are gravity, electromagnetism, the strong force and the weak force. The gravitational force between particles is extremely small and it is negligible compared to the other forces. All twelve fundamental particles interact through the weak force, while only electrically charged particles feel the electromagnetic force. Each force is described by a Quantum Field Theory (QFT) corresponding to the exchange of a spin-1 force-carrying particle, referred to as a gauge boson. The massless photon is the gauge boson responsible for the electromagnetic interaction. The strong force is mediated by the gluon, which is massless like the photon. Only quarks can interact through the strong force, but not leptons. Weak interactions are mediated by massive W and Z bosons. The gauge bosons are summarised in Table 2.2.

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

Table 2.1: The twelve fundamental fermions in the Standard Model, divided into three generations of quarks and leptons. The electric charge and the current mass of each particle are shown [5].

Force	Boson	Spin	Electric charge	Mass (GeV)
Strong	Gluon (g)	1	0	0
Electromagnetism	Photon (γ)	1	0	0
Weak	W boson (W^\pm)	1	± 1	80.4
	Z boson (Z)	1	0	91.2

Table 2.2: The three fundamental forces described by the Standard Model, and their corresponding gauge bosons [5].

The final element of the Standard Model is the Higgs boson, which has a mass of 125.11 ± 0.11 GeV [6]. It was discovered by the CMS and ATLAS experiments at the Large Hadron Collider in 2012 [7, 8]. Unlike the other Standard Model particles, it is a scalar particle (spin-0) and is the only fundamental scalar discovered to date. Crucially, the Higgs boson provides the mechanism by which all other particles acquire mass, with the exception of neutrinos for which the mechanism is still unknown. The Higgs boson can be thought of as an excitation of the Higgs field, which is believed to have a non-zero vacuum expectation value. Particles acquire their masses by interacting with the non-zero Higgs field, which is known as the Higgs mechanism. This will be discussed in more detail in Section 2.7.

2.3 Introduction to gauge theories

In QFT, particles are described by excitations of quantum fields which satisfy quantum mechanical field equations. The dynamics of the QFT can be expressed using a Lagrangian density. Spin-0 scalar particles with mass m can be described as excitations of a scalar field $\phi(x^\mu) = \phi(t, x, y, z)$. The Lagrangian for a free non-interacting scalar field can be written as:

$$\mathcal{L}_S = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}m^2 \phi^2, \quad (2.1)$$

where $\partial_\mu \equiv \frac{\partial}{\partial x^\mu}$ represent the derivatives with respect to each of the four space-time coordinates. The first term in Eq. (2.1) represents the kinetic term, while the second term is the potential term. In the Lagrangian formalism, the principle of least action can be used to derive the Euler-Lagrange equation for the field, which can be shown to be:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \quad (2.2)$$

Substituting the Lagrangian in Eq. (2.1) into the Euler-Lagrange equation in Eq. (2.2) gives:

$$\partial_\mu \partial^\mu \phi + m^2 \phi = 0, \quad (2.3)$$

which is known as the Klein-Gordon equation for a free scalar field $\phi(x)$. Spin-half particles can be described by a spinor field $\psi(x)$ with the Lagrangian:

$$\mathcal{L}_D = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi, \quad (2.4)$$

where ψ is a four-component complex spinor, $\bar{\psi}$ is the adjoint spinor and γ^μ are the γ matrices. Deriving the Euler-Lagrange equation for the spinor field ψ gives the Dirac equation:

$$i\gamma^\mu (\partial_\mu \psi) - m\psi = 0, \quad (2.5)$$

which describes the dynamics of spin-half fermions.

2.4 Quantum electrodynamics

Requiring the Dirac equation to be invariant under a U(1) local gauge symmetry introduces the electromagnetic interaction. This can be expressed as the invariance of the Lagrangian under the local phase transformation:

$$\psi(x) \rightarrow \psi'(x) = e^{iq\chi(x)}\psi(x), \quad (2.6)$$

where q is the coupling strength. It is a local transformation as the phase $\chi(x)$ is a function of the space-time coordinates. Under this U(1) local phase transformation, the Lagrangian for a free spin-half particle shown in Eq. (2.4) is not invariant. The required gauge invariance can be restored by replacing the derivative ∂_μ in Eq. (2.4) with the covariant derivative D_μ :

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu, \quad (2.7)$$

where A_μ is a new field. To achieve gauge invariance under a local U(1) gauge transformation, the new field is required to transform as:

$$A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu \chi. \quad (2.8)$$

Hence the introduction of the gauge field A_μ with well-defined gauge transformation properties is necessary for local gauge invariance. The Lagrangian for quantum electrodynamics (QED) which describes the fields for the electron (with $q = -e$) can therefore be written as:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu \partial_\mu - m_e)\psi + e\bar{\psi}\gamma^\mu \psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (2.9)$$

where there is a term describing the interaction of the fermion with the field A_μ , which can be identified as the photon. The kinetic term contains the field $F_{\mu\nu}$, defined as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The kinetic term $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ can be shown to be invariant under U(1) local phase transformations.

An important result, referred to as the Noether's theorem [9], states that there exists a conserved quantity associated with a symmetry of the Lagrangian. The U(1) symmetry in QED leads to the conservation of the electric charge.

2.5 Quantum chromodynamics

For quantum chromodynamics (QCD), which is the QFT of the strong interaction, the underlying symmetry corresponds to invariance under SU(3) local phase transformations:

$$\psi(x) \rightarrow \psi'(x) = \exp [ig_S \boldsymbol{\alpha}(x) \cdot \hat{\mathbf{T}}] \psi(x), \quad (2.10)$$

where $\hat{\mathbf{T}} = \{T^a\}$ are the eight generators of the SU(3) group, which are related to the Gell-Mann matrices by $T^a = \frac{1}{2}\lambda_a$, $\boldsymbol{\alpha}(x)$ are eight functions of the space-time coordinates, and g_S is the coupling strength of the strong interaction. As the generators are represented by 3×3 matrices, three additional degrees of freedom in ψ are introduced, which can be represented by a three component vector. The new degrees of freedom are referred to as colours, which are labelled as red, blue and green. To achieve local gauge invariance, eight new fields $G_\mu^a(x)$ are introduced, where the index $a = 1, \dots, 8$ each corresponds to one of the eight generators of the SU(3) symmetry. Following the same prescription as before to achieve gauge invariance, the partial derivatives ∂_μ are replaced by covariant derivatives D_μ :

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_S T^a G_\mu^a(x). \quad (2.11)$$

The gauge fields $G_\mu^a(x)$ are required to transform under a local SU(3) gauge transformation as:

$$G_\mu^a \rightarrow G_\mu^{a'} = G_\mu^a - \partial_\mu \alpha_a - g_S f_{abc} \alpha_b G_\mu^c. \quad (2.12)$$

The last term in Eq. (2.12) arises due to the fact that the generators of the SU(3) symmetry do not commute, and the f_{abc} are known as the structure constants of the SU(3) group, which are defined by the commutation relations $[\lambda_a, \lambda_b] = 2if_{abc}\lambda_c$. As the generators of SU(3) do not commute, QCD is known as a non-Abelian gauge theory. The fields G_μ^a are the gluons of QCD. The additional term in Eq. (2.12) leads to gluon self-interactions. The Lagrangian for quantum chromodynamics can then be written as:

$$\mathcal{L}_{QCD} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - g_S \bar{\psi}\gamma^\mu T^a G_\mu^a \psi - \frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a, \quad (2.13)$$

where $G_a^{\mu\nu} = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu - g_S f_{abc} G_b^\mu G_c^\nu$ is the field strength tensor of QCD.

There is a lot of experimental evidence for the existence of quarks, for example from the discovery of the Ω baryon (which is formed by three strange quarks) [10] and deep

inelastic scattering experiments [11, 12]. However, quarks have never been observed as free particles. This can be explained by colour confinement, which states that coloured objects are always confined to colour singlet states and that no objects with non-zero colour charge can propagate as free particles [13]. The confinement leads to quarks and gluons always being confined to colourless objects. Therefore, unlike photons, gluons do not propagate over long distances. Moreover, only certain combinations of quarks and antiquarks are allowed to form bound hadronic states because the resulting states are required to be colour singlets. They can be characterised as mesons ($q\bar{q}$), baryons (qqq) and antibaryons ($\bar{q}\bar{q}\bar{q}$). QCD has a property known as asymptotic freedom, which means the strength of the coupling $\alpha_S \sim g_S^2$ decreases with the scale of energy transfer [14, 15]. Therefore perturbation theory can only be used at high energy scales.

The proton not only contains two up quarks and a down quark (known as the valence quarks), but also contains a sea of virtual gluons that give rise to quarks (known as sea quarks) through pair production. The quarks inside the proton interact with each other through the exchange of gluons. This results in a distribution of quark momenta within the proton. These distributions are expressed in terms of Parton Distribution Functions (PDFs). They are extracted from high-energy measurements in different experiments.

2.6 Electroweak sector

The charged-current weak interaction is associated with invariance under SU(2) local phase transformations:

$$\phi(x) \rightarrow \phi'(x) = \exp[i g_W \boldsymbol{\alpha}(x) \cdot \mathbf{T}] \phi(x), \quad (2.14)$$

where \mathbf{T} are the three generators of the SU(2) group, $\boldsymbol{\alpha}(x)$ are three functions which specify the local phase and g_W is the coupling strength of the weak interaction. The generators can be written in terms of the Pauli spin matrices as $\mathbf{T} = \frac{1}{2} \boldsymbol{\sigma}$. Three gauge fields W_μ^k ($W_\mu^{(1)}, W_\mu^{(2)}$ and $W_\mu^{(3)}$) are introduced to achieve local gauge invariance. The physical W bosons can be identified as linear combinations of the gauge fields:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^{(1)} \mp i W_\mu^{(2)} \right) \quad (2.15)$$

As it was observed that the weak charge-current interaction couples only to left-handed (LH) chiral particle states and right-handed (RH) chiral antiparticle states [16], the symmetry group of the weak interaction is referred to as SU(2)_L.

One of the main goals of particle physics is to provide a unified description of the fun-

damental particles and their interactions. In the 1960s, Glashow, Salam and Weinberg (GSW) developed a unified picture of the electromagnetic and weak interactions [17, 18, 19]. In the GSW electroweak model, the $U(1)$ gauge symmetry of electromagnetism is replaced by a new $U(1)_Y$ local gauge symmetry:

$$\psi(x) \rightarrow \psi'(x) = \exp \left[ig' \frac{Y}{2} \zeta(x) \right] \psi(x), \quad (2.16)$$

which gives rise to a new gauge field B_μ that couples to a new charge referred to as hypercharge Y . The coupling strength is denoted by g' and the local phase is given by $\zeta(x)$. In the unified electroweak model, the photon and the Z boson are written as linear combinations of the B_μ and the neutral $W_\mu^{(3)}$ of the weak interaction:

$$A_\mu = B_\mu \cos \theta_W + W_\mu^{(3)} \sin \theta_W, \quad (2.17)$$

$$Z_\mu = -B_\mu \sin \theta_W + W_\mu^{(3)} \cos \theta_W, \quad (2.18)$$

where θ_W is the weak mixing angle. The underlying gauge symmetry is the $U(1)_Y$ of weak hypercharge and the $SU(2)_L$ of the weak interaction, which can be written as $SU(2)_L \times U(1)_Y$. For invariance under transformations of both gauge groups, it can be shown that

$$q = I_W^{(3)} + \frac{1}{2}Y \quad (2.19)$$

and

$$e = g_W \sin \theta_W = g' \cos \theta_W, \quad (2.20)$$

where $I_W^{(3)}$ is the third component of the weak isospin, which is a quantum number of the weak interaction.

The Lagrangian for the electroweak theory can be written as:

$$\mathcal{L}_{EW} = i\bar{\psi}\gamma^\mu D_\mu \psi - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a, \quad (2.21)$$

where $D_\mu = \partial_\mu + \frac{1}{2}ig'YB_\mu + ig_W\mathbf{T} \cdot \mathbf{W}_\mu$, and $B_{\mu\nu}$ and $W_{\mu\nu}$ are the field strength tensors for B_μ and W_μ^a respectively.

2.7 The Higgs mechanism

In the Standard Model, particles acquire masses through interactions with the Higgs field, as the inclusion of particle masses in the Lagrangian breaks local gauge invariance. The

Higgs mechanism [20, 21, 22, 23, 24, 25] is based on spontaneous symmetry breaking of the $SU(2)_L \times U(1)_Y$ gauge symmetry. A complex scalar $SU(2)$ doublet ϕ is introduced:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \quad (2.22)$$

where ϕ^+ and ϕ^0 are complex scalar fields. The Lagrangian for the complex scalar doublet is given by:

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \quad (2.23)$$

with the Higgs potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (2.24)$$

The covariant derivatives in the electroweak theory (as shown in Section 2.6) are used in the Lagrangian such that it respects the $SU(2)_L \times U(1)_Y$ local gauge symmetry. For $\mu^2 < 0$, the Higgs potential has an infinite set of degenerate minima which satisfy:

$$\phi^\dagger \phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda}, \quad (2.25)$$

where v denotes the vacuum expectation value of the Higgs field. The choice of the physical vacuum state spontaneously breaks the symmetry of the Lagrangian. After symmetry breaking, the neutral photon is required to remain massless, so the minimum of the potential corresponds to a non-zero vacuum expectation value only for the neutral scalar field ϕ^0 :

$$\langle 0 | \phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (2.26)$$

The fields can then be expanded about the minimum. In the unitary gauge, the Higgs doublet can be written as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (2.27)$$

The resulting Lagrangian is referred to as the Salam-Weinberg model. The lower component of the Higgs doublet is neutral and has $I_W^3 = -\frac{1}{2}$. With Eq. (2.19), it can be seen that the Higgs doublet has hypercharge $Y = 1$. Therefore the covariant derivative acting

on the Higgs doublet ϕ can be written as :

$$D_\mu \phi = \frac{1}{2} [2\partial_\mu + (ig_W \boldsymbol{\sigma} \cdot \mathbf{W}_\mu + ig' B_\mu)] \phi \quad (2.28)$$

The term in the Lagrangian that generates the masses of the gauge bosons is $(D_\mu \phi)^\dagger (D^\mu \phi)$, which is given by:

$$\begin{aligned} (D_\mu \phi)^\dagger (D^\mu \phi) = & \frac{1}{2} (\partial_\mu h) (\partial^\mu h) + \frac{1}{8} g_W^2 \left(W_\mu^{(1)} + i W_\mu^{(2)} \right) \left(W^{(1)\mu} - i W^{(2)\mu} \right) (v+h)^2 \\ & + \frac{1}{8} \left(g_W W_\mu^{(3)} - g' B_\mu \right) \left(g_W W^{(3)\mu} - g' B^\mu \right) (v+h)^2 \end{aligned} \quad (2.29)$$

The mass of the W boson can be determined by the quadratic terms in W_μ^1 and W_μ^2 . The terms that are quadratic in the neutral W^3 and B fields can be written in matrix form. Diagonalising the mass matrix and finding the eigenvalues give the masses of the photon and the Z boson. The gauge bosons masses are therefore found as:

$$m_W = \frac{1}{2} g_W v, m_A = 0 \text{ and } m_Z = \frac{1}{2} v \sqrt{g_W^2 + g'^2}. \quad (2.30)$$

A mass term for the Higgs boson arises from the potential, which can be written as $m_H = \sqrt{-2\mu^2}$.

The same mechanism can be used to provide mass terms for quarks and charged leptons. The Yukawa interaction term is introduced, which is invariant under $SU(2)_L \times U(1)_Y$ transformations:

$$\mathcal{L}_{\text{Yukawa}} = -g_f \left(\bar{L} \phi R + \bar{R} \phi^\dagger L \right), \quad (2.31)$$

where g_f corresponds to the Yukawa coupling, L represents a $SU(2)$ doublet of left-handed chiral fermions, and R denotes a $SU(2)$ singlet of right-handed fermions. After spontaneous symmetry breaking, the fermion mass is proportional to the Yukawa coupling, given by $m_f = v g_f / \sqrt{2}$. The masses of fermions are free parameters of the Standard Model and must be measured experimentally.

The right-handed chiral neutrino states do not participate in any of the interactions of the Standard Model, so there is no direct evidence that they exist. However studies of neutrino oscillations show that neutrinos have mass, so there must be a corresponding mass term in the Lagrangian. The exact mechanism under which neutrinos obtain masses is still an open question in particle physics.

Chapter 3

Dark matter and Hidden Valley models

3.1 Introduction

The Standard Model has made many successful predictions, however there remain many open questions to be answered. One of the biggest questions is the nature of dark matter (DM). From various astrophysical observations, the existence of a “dark” form of matter is inferred, which is matter that does not interact with the electromagnetic force but has mass. Aside from the measurement of galaxy rotation curves, there is also other evidence of dark matter, for example from gravitational lensing effects (with the Bullet cluster being the most promising example [26]) and the cosmic microwave background [2]. About 27% of the energy content of our Universe is believed to be made up of dark matter, compared to only 5% being visible matter [1, 2, 3, 4].

Many theories beyond the Standard Model (BSM) predict the existence of a dark sector which can be probed with collider experiments. They often involve a dark matter candidate and a mediator particle that couples to both SM and dark matter particles. In this Chapter, Hidden Valley models will be discussed, which are the target of the search.

3.2 Hidden Valley models

One class of BSM models that could explain dark matter is known as Hidden Valley models, in which the SM gauge group is extended by a new, non-Abelian gauge group [27, 28, 29, 30, 31, 32]. Dark matter particles are charged under this new gauge group (the dark gauge group). Hidden Valley models can also be ingredients in models that address different problems in particle physics, such as the stability of the electroweak hierarchy [33, 34, 35] (which refers to the large difference between the electroweak scale and the Planck scale), the matter-antimatter asymmetry [36], and the origin of neutrino masses

[37, 38].

Hidden Valley models can produce events with a high multiplicity of non-isolated dark particles, known as dark showers. In confining Hidden Valley models, the dark particles assemble themselves to form dark hadrons. Dark hadrons can possibly decay into SM particles with observable lifetimes [39, 40, 41, 42], creating displaced signatures in detectors. These signatures are displaced with respect to the proton-proton collision point, so they would appear to be coming from nowhere. Dark showers containing sprays of dark hadrons are expected to create high-multiplicity displaced signatures [43, 44, 45, 46, 47, 48, 49, 50].

3.2.1 Vector portal benchmark model

The Hidden Valley module [51, 52] in PYTHIA 8 [53] was used to generate benchmark models which describe perturbative dark parton showers. The benchmark models are described by a SU(3) gauge theory with one dark sector flavour. The models take advantage of the good properties of theories in the QCD-like regime. These theories have small 't Hooft couplings $\lambda \equiv g^2 N_c$, where g and N_c are the gauge coupling and the number of colours at the energy scale Q at which the dark shower is produced. This allows good modelling of parton showering using perturbation theory.

The production of dark showers through the decay of the SM Higgs is considered. Currently upper limits are set at about 10% on the branching ratio of Higgs to invisible decays [54, 55], which motivate the search for dark shower through Higgs decay. Following production and showering, hadronization is described by a Lund string model in the PYTHIA Hidden Valley module [56] which leads to a simplified hadron sector. It only includes spin-zero and spin-one mesons, but not baryons and excited mesons.

Models with one dark flavour are first considered, which correspond to a hadron spectrum that consists of a spin-zero meson $\tilde{\eta}$, and a spin-one meson $\tilde{\omega}$, both carrying no flavour symmetry. The mesons are denoted in loose analogy with their SM counterparts. It is assumed that if the $\tilde{\omega} \rightarrow \tilde{\eta}\tilde{\eta}$ channel is kinematically open, the decay occurs promptly with 100% branching ratio so that the $\tilde{\eta}$ multiplicity in the model could be varied by changing the mass ratio between the dark mesons [41]. Figure 3.1 shows a schematic of how the dark mesons are formed starting from the production of dark partons. The Higgs boson first decays into a pair of dark partons $\psi\bar{\psi}$, which then hadronise through a dark analog to quantum chromodynamics into dark mesons. The vector portal model is studied, where the dark vector meson has a displaced decay into Standard Model particles, with a high branching ratio for the decay into muons.

There are a few independent parameters in the Hidden Valley models - the mass of the

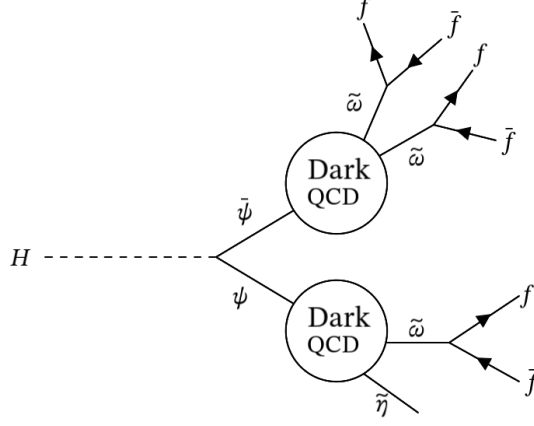


Figure 3.1: Feynman diagram for the vector portal model. Dark partons $\psi\bar{\psi}$ are first produced from the decay of the SM Higgs boson, which then hadronise to form dark vector mesons $\tilde{\omega}$ and dark scalar mesons $\tilde{\eta}$. The $\tilde{\omega}$ then undergo displaced decay into SM fermions, f .

spin-zero meson $m_{\tilde{\eta}}$, the mass of the spin-one meson $m_{\tilde{\omega}}$, and the dark sector confinement scale $\tilde{\Lambda}$ [57]. These are represented by the parameters $m_{\tilde{\eta}}$, $\xi_{\omega} \equiv m_{\tilde{\omega}}/m_{\tilde{\eta}}$ and $\xi_{\Lambda} \equiv \tilde{\Lambda}/m_{\tilde{\omega}}$. The dark quark mass is set to be approximately half the mass of the $\tilde{\eta}$ by choice [52]. As the $\tilde{\omega}$ has a higher spin than the $\tilde{\eta}$, it is theoretically more probable to have $m_{\tilde{\omega}} \approx \tilde{\Lambda} \gtrsim m_{\tilde{\eta}}$. In the benchmark models it is assumed that $m_{\tilde{\omega}} = \tilde{\Lambda} = m_{\tilde{\eta}}$, such that the decay of $\tilde{\omega}$ into $\tilde{\eta}$ is kinematically closed to prevent $\tilde{\omega}$ from strongly decaying into other dark sector states [41].

It is also assumed that only one of the dark mesons ($\tilde{\eta}$ or $\tilde{\omega}$) has a detector-relevant lifetime. The unstable dark meson is called the visibly-decaying dark particle (VDP). Other dark hadrons are assumed to either decay promptly to the VDP, or escape the detector as missing energy. Even though this choice does not account for the full range of possible dark shower topologies, it still encompasses the phenomenology of a wide class of models, and describes a wide range of distinct signatures with a minimal number of arbitrary parameters [41].

A few theory priors are imposed when constructing the benchmark models. It is assumed that there are no new sources of SM flavour violation. Operators up to dimension five are allowed in the infrared effective theory. In the vector portal, the dark vector meson decays into SM particles through a coupling to the SM electromagnetic current [27]. A kinetically mixed dark photon mediator is incorporated, which leads to leptonic and hadronic decays with comparable branching ratios. This is realised by considering an elementary vector boson A' which mixes with the SM hypercharge field strength in the UV-completed theory:

$$\mathcal{L}_{UV} \supset g A'^{\mu} \bar{\psi} \gamma_{\mu} \psi + \frac{\epsilon}{2} B^{\mu\nu} F'_{\mu\nu}, \quad (3.1)$$

where g is the coupling to dark sector quarks and ε is the mixing parameter with the SM hypercharge. This leads to the following in the infrared effective theory (as long as $m_{A'} > m_{\tilde{\omega}}$):

$$\mathcal{L}_{IR} \supset \varepsilon_{\text{eff}} e J_{EM}^\mu \tilde{\omega}_\mu, \text{ where } \varepsilon_{\text{eff}} \approx g \frac{f_{\tilde{\omega}} m_{\tilde{\omega}}^2}{m_{A'}^2} \varepsilon, \quad (3.2)$$

and $f_{\tilde{\omega}}$ is the dimensionless decay constant of the vector meson $\tilde{\omega}$, taken to be $f_{\tilde{\omega}} = 1$ as a conservative estimate [58]. The electromagnetic current is given by $J_{EM}^\mu = q_f \bar{\psi} \gamma^\mu \psi$, where q_f is the charge of the SM fermion. Therefore Eq. (3.2) implies that $\tilde{\omega}$ decays exclusively to charged final SM states. If $m_{\tilde{\omega}} \gtrsim 2$ GeV, its width is given by

$$\Gamma_{\tilde{\omega}} = \sum_f \frac{\alpha_{EM}}{3} q_f^2 C_f \varepsilon_{\text{eff}}^2 m_{\tilde{\omega}} \left(1 + 2 \frac{m_f^2}{m_{\tilde{\omega}}^2} \right) \sqrt{1 - \frac{4m_f^2}{m_{\tilde{\omega}}^2}}, \quad (3.3)$$

where α_{EM} is the electromagnetic fine structure constant, m_f is the mass of the SM fermion f , and $C_f = 1$ ($C_f = 3$) for leptons (quarks). If $m_{\tilde{\omega}} \lesssim 2$ GeV, there are large corrections to the width coming from mixing with SM hadrons. In this case the width is extracted from data in Ref. [59].

The branching ratio for the decay of the dark meson into different SM fermions against the dark meson mass is shown in Fig. 3.2 for the vector portal. It can be seen that the branching ratio for decay into muons is about 10-30%, which is a high branching ratio across the full range of mass. Therefore, the search targets decays into dimuons and uses dimuon invariant mass to extract the signal.

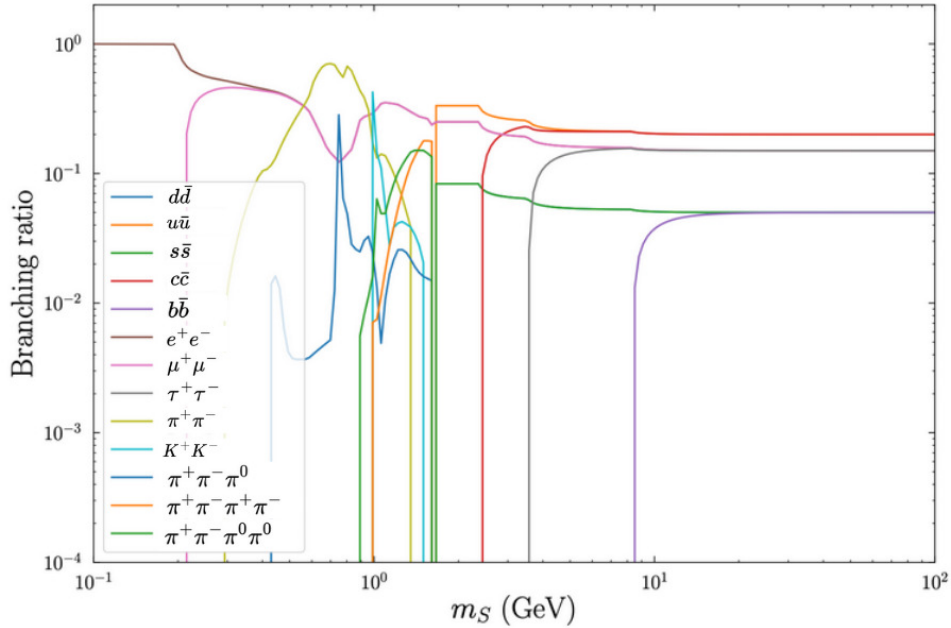


Figure 3.2: Branching ratio for the decay of the dark meson into different SM fermions for the vector portal as a function of the dark meson mass.

A heavy, kinematically-mixed dark photon is used for UV completion, which allows the theory to be well-defined at arbitrarily high energies. The dark photon is specified by its mass $m_{A'}$, its coupling to the dark sector quarks g and the mixing parameter with the SM hypercharge ε . The mixing parameter is constrained by searches for low and high mass dilepton resonances [60, 61, 62], as well as electroweak precision tests (EWPT) [63]. With a single flavour of dark sector fermions, $N_c = 3$ dark colours, and assuming $g = 1$, it was found that the EWPT constraints provide the strongest bound in most of the relevant mass range for A' [41]. To minimize ε_{eff} , we choose $m_{A'} = 20$ GeV, which leads to the following approximate lower bound on the $\tilde{\omega}$ lifetime [41]:

$$c\tau \gtrsim 7 \times 10^{-3} \text{ cm} \times \left(\frac{1 \text{ GeV}}{m_{\tilde{\omega}}} \right)^5 \quad (3.4)$$

In this decay portal, the spin-zero meson $\tilde{\eta}$ is stable so it escapes the detector as missing energy. This implies that a semi-visible jet phenomenology [64, 65] is expected in the regime where the spin-one meson $\tilde{\omega}$ decays promptly. The semi-visible jet signature is defined by significant missing transverse energy along the direction of the jet. The visible hadronic states from the decay of $\tilde{\omega}$ would result in jets, while the $\tilde{\eta}$ meson is stable and remains invisible, leading to missing transverse momentum.

3.2.2 Extended benchmark models

Extended Hidden Valley models with more complex decay topologies are also studied [66]. The dark shower is again assumed to be initiated through the decay of the SM Higgs boson, as this is theoretically the simplest production mode without needing to introduce a new, heavy mediator. The SM Higgs also has a narrow decay width, which makes it a sensitive probe for Hidden Valley models [67].

In the extended benchmark models there are two dark flavours instead of one, which leads to a wider spectrum of dark hadrons, including three dark pions (π_1, π_2 and π_3) and a heavier pseudoscalar η . A massive U(1) interaction is included to facilitate different decay chains. This U(1) is spontaneously broken, resulting in a massive dark photon A' that kinetically mixes with the SM photon to provide a muon-philic decay portal back to the SM. Both the π_3 and η can decay into a pair of A' through the dark sector chiral anomaly if this channel is kinematically open. On the other hand, the π_1 and π_2 are stable on collider timescales and they contribute to the missing transverse energy of the event.

Two scenarios commonly considered in the literature and denoted scenario A and B1 are studied [66]. The Feynman diagrams for the decays in these scenarios are shown in Fig. 3.3. In both scenarios, the pseudoscalar η is assumed to decay promptly into three dark pions. In scenario A, the dark pion π_3 decays promptly into two dark photons, which

can then undergo displaced decays into muons given the mixing angle between the dark photon and the SM photon is small. The vector sum of the momenta of the muons in the dimuon vertex points back to the beamline, so this signature is referred to as a pointing scenario. In scenario B1, the dark gauge coupling is chosen to be small such that the decay of the dark pion π_3 into dark photons becomes displaced. The dark photons can then decay into muons, and the dark photon lifetime is taken to be prompt in this scenario. As a result, the dimuon vertices overlap each other and the momentum vectors of the dimuon vertices do not point back to the beamline. This is referred to as a non-pointing scenario. The analysis is aimed to be sensitive to both the pointing and the non-pointing scenarios.

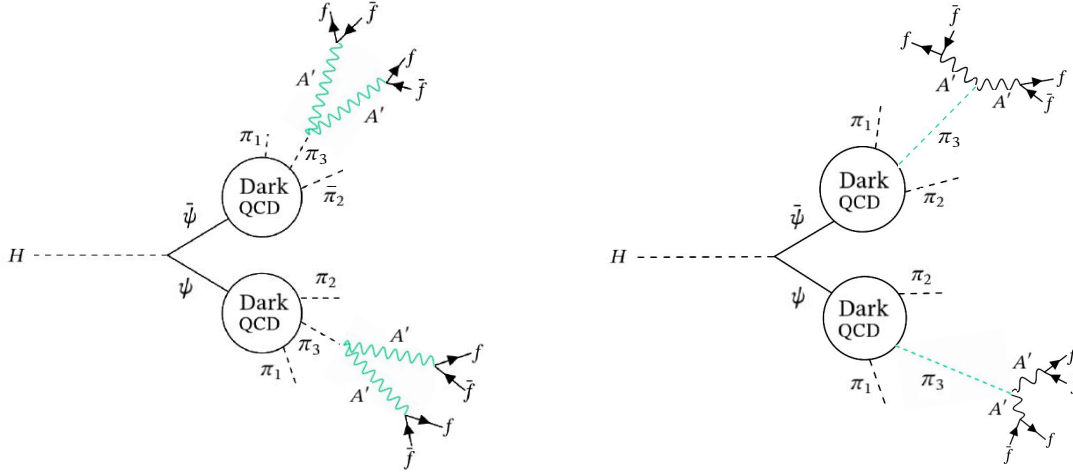


Figure 3.3: Feynman diagram for the Scenario A model (left) and the Scenario B1 model (right). In these extended models the dark hadronization produces a spectrum of dark mesons, including the dark pions π_1, π_2 and π_3 . The π_3 then decays into SM fermions through the dark photon A' . The A' is a long-lived particle in Scenario A, while the π_3 is a long-lived particle in Scenario B1, as indicated by the green lines.

For scenario A and scenario B1, it is assumed that $m_\eta = \tilde{\Lambda}$. There is a mixing angle θ that parametrizes the isospin violation in the coupling of the dark photon to dark sector quarks. The mixing angle is taken to be $\sin \theta = 0.1$ so that the isospin breaking is relatively small, which results in the dark pions having similar masses, i.e. $m_{\pi_1} \approx m_{\pi_2} \approx m_{\pi_3}$. This gives a simpler mass spectrum, and also ensures the simulation from PYTHIA provides a reasonable approximation to the hadronization process [66]. The branching ratio $\mathcal{B}(\pi_3 \rightarrow A'A')$ is assumed to be one.

Figure 3.4 shows the multiplicity of dark mesons for representative mass models of the vector portal (left), and scenario A and B1 (right). The multiplicity increases with decreasing mass of the dark meson, as expected since less energy is required to produce a lighter meson. The multiplicity is found to be very similar between scenario A and scenario B1 for the same meson mass spectrum, as the two models have similar decay processes, with the only difference being whether the dark pion π_3 or the dark photon is

long-lived.

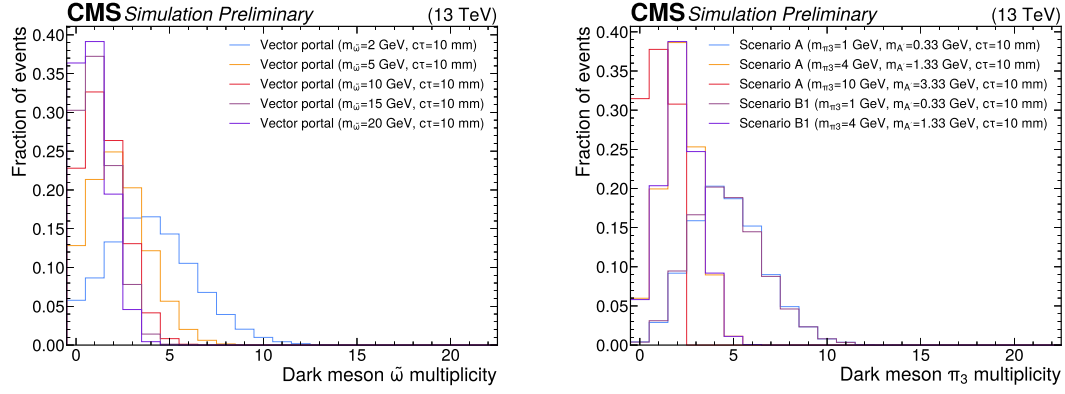


Figure 3.4: The multiplicity of dark vector mesons $\tilde{\omega}$ for representative vector portal models (left), and the multiplicity of dark mesons π_3 for representative scenario A and B1 models (right). The fraction of generated events is shown against the multiplicity of dark mesons.

Chapter 4

The LHC and the CMS experiment

4.1 Introduction

The analysis presented in this thesis uses proton-proton collision data collected with the CMS experiment in 2018 at a centre of mass energy of 13 TeV. In this Chapter, a brief overview of the European Organisation for Nuclear Research (CERN) accelerator complex and the Large Hadron Collider is given. The subdetectors of the CMS experiment are then discussed, which include the tracker, the electromagnetic calorimeter, the hadron calorimeter and the muon chambers. Finally, the data acquisition and trigger system are discussed.

4.2 The Large Hadron Collider

The Large Hadron Collider (LHC) is a hadron accelerator and collider installed in a 26.7 km tunnel between 45 m and 170 m underground [68]. It contains superconducting magnets and has a two-ring structure. Beams of protons or heavy ions (such as lead ions) are collided together at energies of up to 7 TeV or 2.76 TeV per nucleon respectively. There are different experiments that use detectors to analyse the particles that are produced by the collisions, including four detectors that reside at beam intersection points. The largest experiments are the CMS [69] and ATLAS [70] experiments, which use general-purpose detectors to investigate a large range of physics, including searching for physics beyond the Standard Model. The ALICE [71] and LHCb [72] experiments have detectors that focus on specialised physics phenomena, which relate to heavy ion and heavy flavour physics respectively.

The particles are accelerated with the CERN accelerator complex, which is depicted in Fig. 4.1. The acceleration is done in stages. For the acceleration of protons, proton

bunches are first produced by stripping off electrons from hydrogen atoms in hydrogen gas using an electric field. The protons enter the Linear Accelerator 2 (LINAC 2) where they are accelerated to 50 MeV. The protons are then injected into the Proton Synchrotron Booster (PSB), which further raises the energy of the protons to 1.4 GeV. The next two stages are the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), which increase the beam energy to 25 GeV and 450 GeV respectively. Finally the beams reach the LHC, which has a maximum design beam energy of 7 TeV.

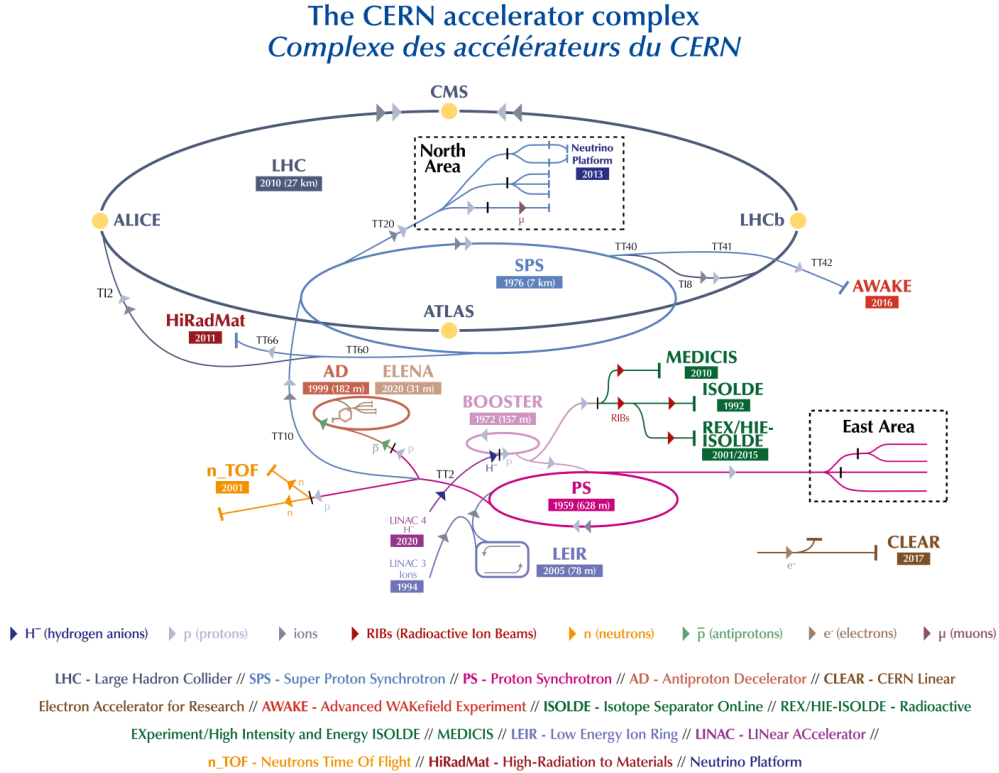


Figure 4.1: A schematic depiction of the CERN accelerator complex. Image credit to CERN [73].

Each of the two proton beams consists of 2808 bunches of protons (separated by 25 ns or around 7.5 m), and each proton bunch contains up to 10^{11} protons by design. The beams are steered by superconducting electromagnets, which provide magnetic fields of up to 8 T. The collision between the proton bunches is referred to as a “bunch crossing”. The resulting interactions in a collision are referred to as an “event”. The number of events produced per second is given by $L\sigma$, where L is the instantaneous luminosity and σ is the cross-section for a particular process. The instantaneous luminosity can be thought of as a measure of how tightly particles are packed inside the proton beams (which depends on the number of protons and the number of bunches), but it also depends on other beam quantities such as the crossing angle. The cross-section describes the probability that two particles will collide and interact in a certain way. The integrated luminosity \mathcal{L} measures

the total number of events during a data-taking period, which is given by $\mathcal{L}\sigma$. When two proton bunches collide, multiple proton-proton collisions can occur simultaneously. Pile-up is defined as the average number of particle interactions per bunch crossing.

4.3 LHC performance

There have been three periods of data taking at the LHC, in which the most recent run is ongoing. The Run 1 data taking spanned the period of 2009 to 2013. After a year long delay due to an unexpected magnet quench, a beam energy of 3.5 TeV was reached in 2010, and it was further increased to 4 TeV in 2012. About 27 fb^{-1} of data were collected. A remarkable achievement during Run 1 was the discovery of the Higgs boson in July 2012. To prepare for the next stage of data taking, the LHC went into the first shutdown period in 2013.

The LHC started its operation again in 2015 and lasted until 2018, which is referred to as Run 2. A beam energy of 6.5 TeV was achieved, which corresponds to a centre of mass energy of $\sqrt{s} = 13 \text{ TeV}$. During Run 2, the LHC achieved an instantaneous luminosity of $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, which was twice its design value. The integrated luminosity collected in the Run 2 proton-proton collision data was about 137 fb^{-1} .

The luminosity delivered to CMS is measured with different systems. Some systems, like the Pixel Luminosity Telescope and the hadronic forward calorimeter, use a separate data acquisition system which operates independently of the main CMS readout. Other methods such as the drift tube luminosity and the pixel cluster counting method use data from existing parts of the CMS detector to perform a measurement using the main CMS data acquisition system. The luminosity measured has a typical precision of about 2%. For example, the systematic uncertainty on the luminosity measurement in 2018 was found to be 2.5% [74].

After more than three years of upgrade and maintenance work following Run 2, the Run 3 period of data taking began in 2022, and it continues until 2026. The beam energy reaches 6.8 TeV, which gives a record collision energy of 13.6 TeV. About 180 fb^{-1} of data have been recorded from 2022 to 2024, which was already more than the data collected in the previous two runs combined. Throughout the different runs, the pileup has gradually increased and exceeded the design value. For example the average pileup in the 2024 data taking was 57, compared to the design value of 20. Both the luminosity and pileup are expected to further increase in the High-Luminosity LHC (HL-LHC), which is a new phase of the LHC starting in 2030. This creates a challenging detector environment, and therefore upgrades will be performed during the shutdown period starting in 2026. The luminosity of the LHC together with the projected luminosity of the HL-LHC are shown

in Fig. 4.2.

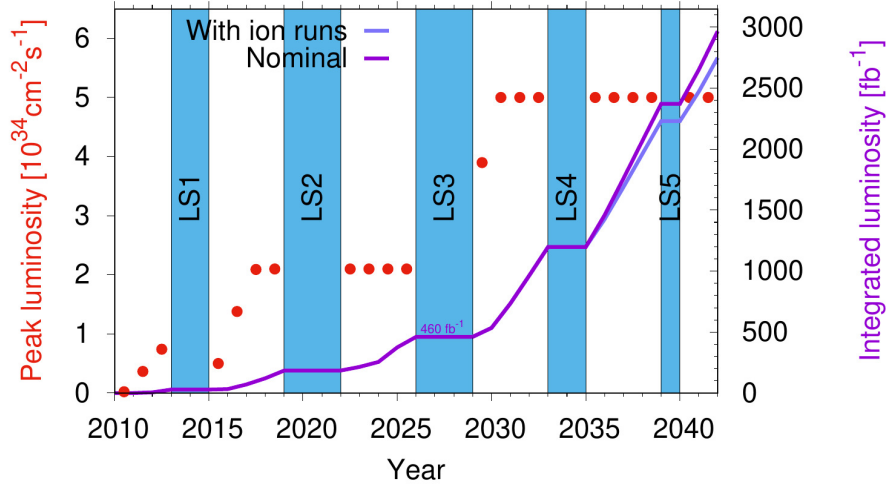


Figure 4.2: The proton-proton luminosity of the LHC, with projections for the HL-LHC up until 2041. Image credit to CERN [75].

4.4 The CMS experiment

The Compact Muon Solenoid (CMS) [69] is one of the two general-purpose detectors at the LHC, which was designed to operate with proton-proton and ion collisions at high energies. The central feature of CMS is a large superconducting solenoid with an internal diameter of 6 m, which provides a magnetic field of 3.8 T over a radius of 3.15 m [76]. The CMS experiment is a cylindrical, layered detector which comprises of multiple subsystems used for detecting and reconstructing particles. A schematic of the detector is shown in Fig. 4.3. The subdetectors at increasing distance from the beamline are the tracker, the electromagnetic calorimeter, the hadron calorimeter and the muon chambers [69].

The coordinate system used to describe the geometry of the CMS experiment has the origin located at the interaction point, as shown in Fig. 4.4. The z -axis points in the beam (longitudinal) direction, the x -axis points towards the centre of the LHC ring, and the y -axis points upwards. In polar coordinates, the azimuthal angle in the transverse plane starting from the x -axis is denoted by ϕ , while the polar angle starting from the z -axis is denoted by θ . To describe the angle between the z -axis and particles in the collider which are highly relativistic, it is convenient to define the pseudorapidity as $\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]$ because differences in η are approximately Lorentz invariant.

The CMS detector has full overage in ϕ but not in η , where it is restricted to $0 < |\eta| < 5$. In the region close to the beam axis (corresponding to high values of η), there is high radiation and the presence of the beam pipe, which make it difficult to measure particles.

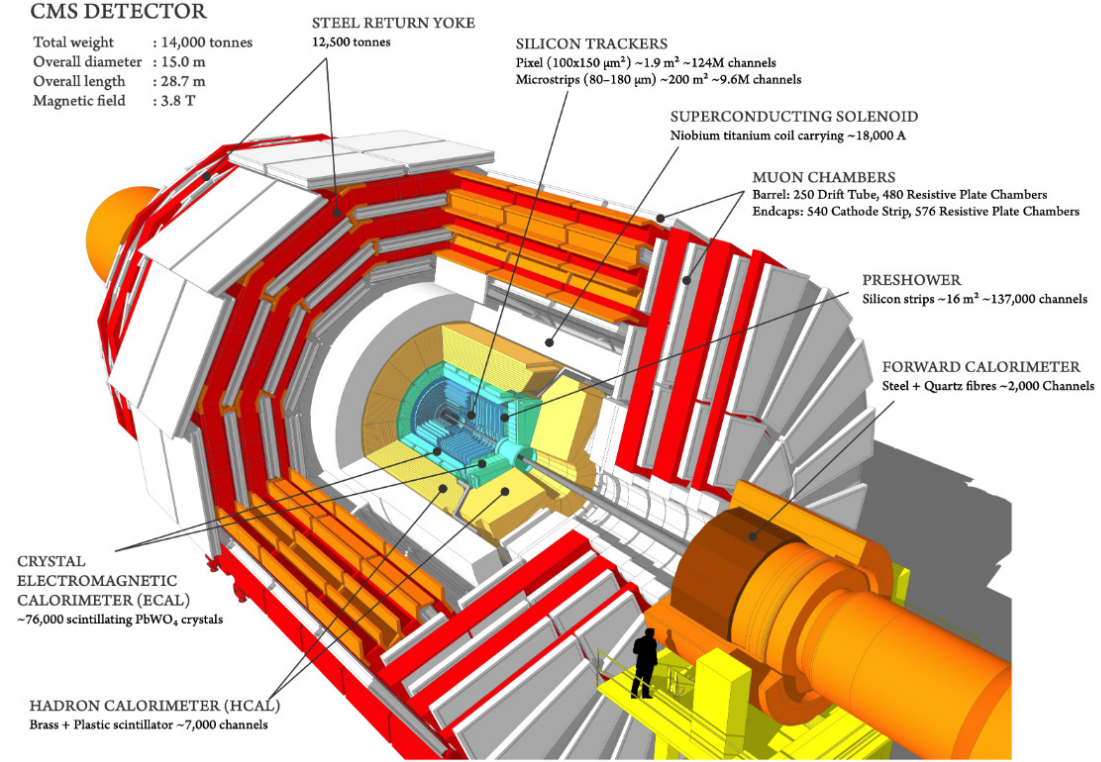


Figure 4.3: A schematic of the CMS detector [77].

In the following subsections, the subdetectors and the data acquisition system will be discussed.

4.4.1 Tracker

The tracking detector [79] is located close to the beamline and is used for measuring the trajectories (referred to as “tracks”) of charged particles which are bent in the magnetic field, and also for estimating the points of origin (referred to as “vertices”) of the particle tracks. It is designed to provide good momentum resolution for high energy (up to $\sim 1000 \text{ GeV}$) particles, as well as good reconstruction efficiency for low-energy ($\sim 1 \text{ GeV}$) tracks and vertex resolution of $\sim 10 \mu\text{m}$.

The tracker is a cylindrical silicon detector which is made up of the pixel detector (inner layer) and the strip detector surrounding it, as shown in Fig. 4.5. It measures the momenta of charged particles in the pseudorapidity range of $|\eta| < 2.5$. The tracker system has a length of 5.8 m and a diameter of 2.5 m, covering a surface area of 200 m^2 . The basic working principle of the silicon devices involves applying a reverse bias voltage to a p-n junction to create a depletion zone free of charge carriers. When a charged particle passes, it induces electron-hole pairs in an electric field, leading to a current pulse to be produced as a “hit” signal. This forms the basis of reconstructing particle tracks and vertices.

The pixel detector is located closest to the interaction point, extending up to $r = 20 \text{ cm}$.

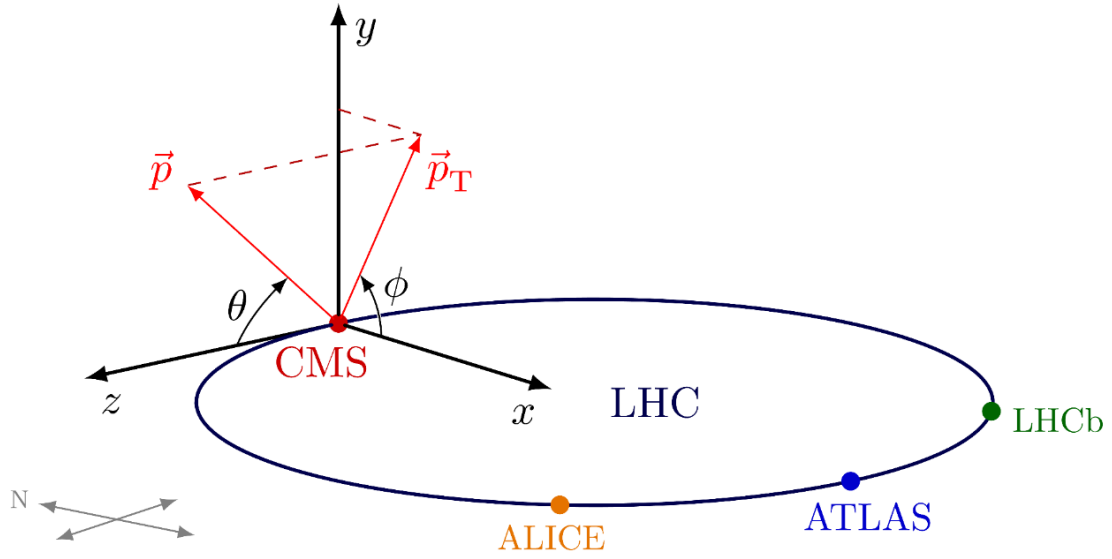


Figure 4.4: Graphical representation of the coordinate system used for the CMS experiment [78].

It performs three-dimensional hit position measurements, with a resolution of $10 \mu\text{m}$ and $30 \mu\text{m}$ in the transverse and the longitudinal plane respectively [80]. The detector is used to ensure good vertex resolution close to the interaction point. It was upgraded to include four barrel layers and three endcap disks during 2016-2017, which led to significant improvement in track reconstruction efficiency and resolution performance. The four concentric barrel layers are at radii of 29, 68, 109, and 160 mm. The pixel detector contains 1856 segmented silicon sensor modules over an area of 1.9 m^2 , resulting in 124 million readout channels in total [81]. It is an important subdetector for the reconstruction of secondary vertices, which are vertices outside the beam profile, which may originate from the decays of unstable particles with different lifetimes [82]. Therefore the reconstruction of secondary vertices is crucial for searches for displaced signatures.

The strip detector is located at $r > 20 \text{ cm}$, covering an area of 198 m^2 . Some of the modules provide two-dimensional hits, while others allow the reconstruction of three-dimensional hit positions. Each of the latter consists of two back-to-back strip modules, in which one module is rotated through a “stereo” angle. Overall, the strip modules are divided into four regions. The two inner regions are referred to as the tracker inner barrel (TIB) and the tracker inner disk (TID), which provide position measurements in the r - ϕ plane with a resolution of 13 - $38 \mu\text{m}$. The inner regions are surrounded by tracker outer barrel (TOB) and tracker endcaps (TECs), where the resolution degrades to 18 - $47 \mu\text{m}$. There are 9.3 million strips in total in the strip detector.

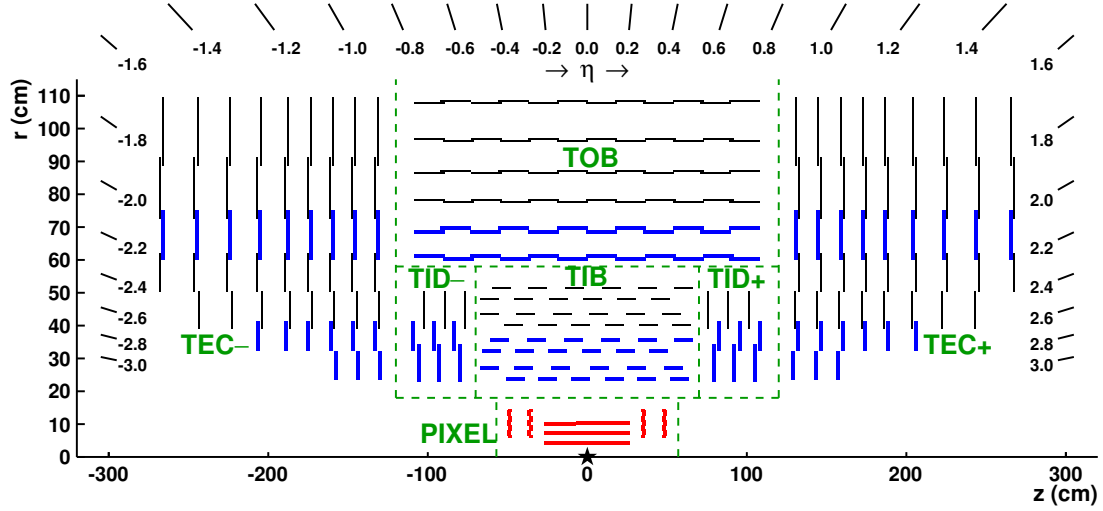


Figure 4.5: A schematic cross-section of the CMS tracker in the r - z plane. The tracker is symmetric about $r = 0$ in this view, so only the top half is shown. The centre of the tracker, which is the approximate position of the proton-proton collision point, is indicated by a star. Strip tracker modules which provide 2-D hits are shown by thin, black lines, and those that permit the reconstruction of 3-D hit positions are shown by thick, blue lines. The pixel modules are shown by red lines and they also provide 3-D hits. Within a given layer, each module is shifted slightly in r or z with respect to its neighbouring modules, which allows them to overlap and avoids gaps in the acceptance [80].

4.4.2 Electromagnetic Calorimeter

The subsystem particles reach after the tracker is the Electromagnetic Calorimeter (ECAL) [83], which performs high precision measurements of the positions and energies of photons and electrons by measuring the properties of the electromagnetic showers they induce. It is made of lead tungstate (PbWO_4) crystals, with a cylindrical electromagnetic barrel (EB) region ($|\eta| < 1.48$) containing 61200 crystals enclosed by two electromagnetic endcaps (EEs) ($1.48 < |\eta| < 3.0$) containing 7324 crystals each. A schematic of the calorimeter is shown in Fig. 4.6. When electrons pass through the material which has a high atomic number, they emit photons through bremsstrahlung radiation. The photons can then produce electrons and positrons through pair production. This forms an electromagnetic shower, which is used to measure the energies of electrons and photons by using the scintillation light produced by the ECAL crystals as the particles travel through them.

Lead tungstate is a very dense ($\rho = 8.3 \text{ gcm}^{-3}$) and radiation-hard material with a short radiation length ($X_0 = 0.89 \text{ cm}$), which make it a good candidate for scintillating crystals. As the ECAL is about $25X_0$ deep, it absorbs almost all of the electromagnetic showers which contain more than 98% of energy for electrons and photons with energies up to 1 TeV. Lead tungstate also has a small Molière radius of 2.2 cm, which allows for compact design of the detector. With compact crystals, it is possible to resolve energy deposits from different particles that are separated by as little as 5 cm, which helps separate charged

and neutral particles when considering the tracker information. Moreover, lead tungstate has a fast response time of 25 ns, which corresponds to the typical bunch spacing, collecting about 80% of the scintillation light produced. Nevertheless, the light yield of the crystal is relatively low, so the signal needs to be amplified by avalanche photodetectors in the barrel system and vacuum phototriodes in the endcaps.

To account for the reduced granularity at the endcaps, the ECAL is equipped with an electromagnetic pre-shower (ES) detector which covers $1.65 < |\eta| < 2.6$. It helps distinguish photons that come from neutral pion decays ($\pi_0 \rightarrow \gamma\gamma$) from prompt photons.

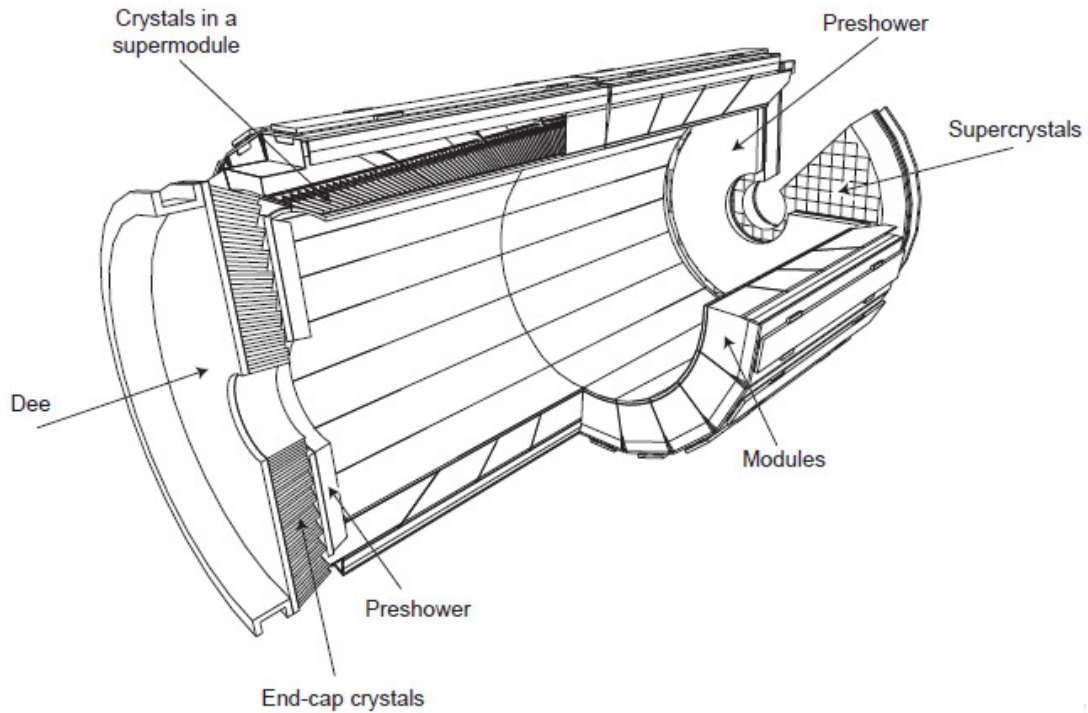


Figure 4.6: A schematic of the CMS Electromagnetic Calorimeter, showing the barrel, endcap and the preshower subsystems. The barrel section comprises of 36 supermodules, each containing 4 modules [84].

4.4.3 Hadron Calorimeter

The Hadron Calorimeter (HCAL) [85] is designed to measure the energies of hadronic showers, which are produced when neutral and charged hadrons interact with the detector material through the strong interaction.

A schematic of the HCAL is shown in Fig. 4.7. It is a sampling calorimeter consisting of alternating layers of non-magnetic brass absorber and plastic scintillator tiles, surrounding the ECAL. The scintillation light is collected by wavelength-shifting fibres, and then translated and amplified by multichannel hybrid photodiodes proportionally to the magnitude of the energy deposits.

The HCAL contains the hadronic barrel (HB) in $|\eta| < 1.4$ and the hadronic endcap (HE) in $1.3 < |\eta| < 3.0$. The HCAL towers have a cross-section of 0.087×0.087 in the η - ϕ space for $|\eta| < 1.6$, and it varies in size with $|\eta|$ after that. Due to space constraints within the magnet cryostat, the HB thickness is limited to 5.8 hadronic interaction lengths (λ_I) at $\eta = 0$, increasing to $10\lambda_I$ at $|\eta| = 1.2$. To extend the size of the HB, the hadronic outer (HO) calorimeter is placed outside the magnet in the region of $|\eta| < 1.26$, which increases the effective thickness of the system to over $10\lambda_I$. Furthermore, hadronic forward (HF) calorimeters are included to cover the region of $3 < |\eta| < 5$ in coarse calorimeter towers, which are designed to endure high radiation levels.

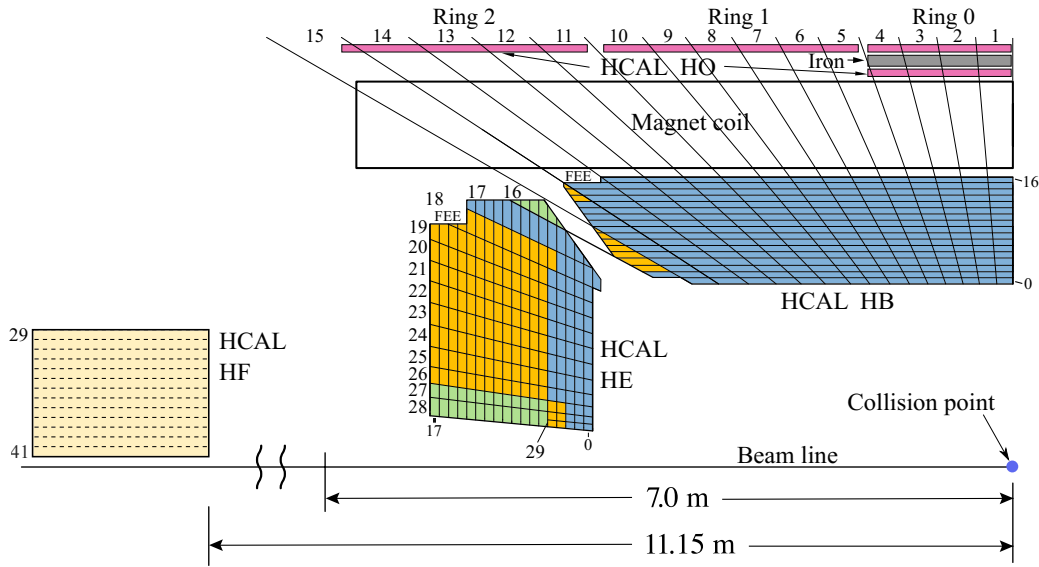


Figure 4.7: A schematic view of one quarter of the CMS HCAL, showing its four major components which are the hadron barrel (HB), the hadron endcap (HE), the hadron outer (HO), and the hadron forward (HF) calorimeters [86].

4.4.4 Muon chambers

Precise muon identification and energy reconstruction are crucial for the CMS detector as muons are key signatures in various Standard Model and beyond the Standard Model processes. The muon system [87] is located outside the magnet cryostat and it covers the region of $|\eta| < 2.4$. It is interleaved with an iron return yoke. A schematic of the muon chambers is shown in Fig. 4.8. The muon system is made up of three types of gas ionisation chambers. As muons pass through the chambers, the gas is ionised, resulting in free electrons which are collected with a strong electric field. This produces a signal in the wires and strips in the chambers.

The drift tube (DT) chambers cover the region of $|\eta| < 1.2$, where the muon rate is relatively low and the magnetic field is weak. With increasing η the muon rate and the mag-

netic field increase, so the cathode strip chambers (CSCs) are used for $0.9 < |\eta| < 2.4$, which have higher radiation hardness and response rate. The muon system can thus be divided into three regions: barrel ($|\eta| < 1.2$), two endcaps ($0.9 < |\eta| < 2.4$) and an overlap region ($0.9 < |\eta| < 1.2$). The resistive plate chambers (RPCs) serve as a complementary system, covering the range of $|\eta| < 1.6$. They offer fast response and are important for associating signals to the correct bunch crossing for triggering, even though they have a much coarser position resolution than DTs or CSCs.

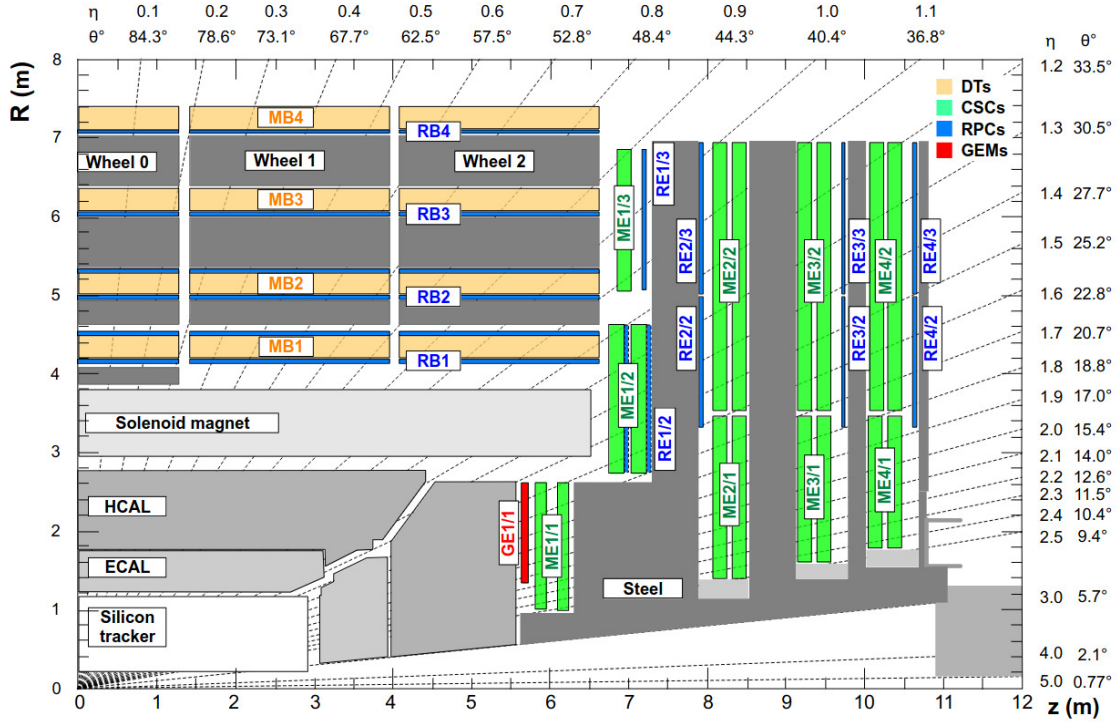


Figure 4.8: A schematic view in the r - z plane of a CMS detector quadrant. The interaction point is located at the lower left corner. The locations of the various muon stations are shown in different colours and labels: drift tubes (DTs) with labels MB, cathode strip chambers (CSCs) with labels ME, resistive place chambers (RPCs) with labels RB and RE, and gas electron multipliers (GEMs) with labels GE. The M denotes muon, B denotes barrel, and E denotes endcap. The magnet yoke is represented by the dark grey areas [88].

4.4.5 Data acquisition and trigger system

With a bunch spacing of 25 ns, collisions happen at the LHC at a rate of 40 MHz. It is impossible to write out and store all the events, and even if it became possible not all events are of interest, for example they might be low-energy inelastic proton-proton scattering rather than hard head-on collisions. Therefore, a trigger system is in place to select events which are then stored for subsequent analysis.

The trigger system at CMS has a two-tier structure [89, 90, 91]. The first level, known as the Level-1 (L1) trigger, consists of custom hardware which accepts events at a rate

of 100 kHz for further processing. The second level, referred to as the high-level trigger (HLT), is made up of a farm of computers which implement software to run high-level physics algorithms, producing a final output rate of about 1 kHz. The events which pass the L1 and HLT decisions are then stored to disk in the data acquisition system.

The L1 trigger is the first level of selection in the system [92]. It is used for making a fast decision on whether to keep an event, with a latency of less than $4\ \mu\text{s}$. For example, it discards simple QCD-induced multijet events (which are abundantly produced), while maintaining good efficiency for events which are useful for Standard Model precision measurements and searches for beyond the Standard Model physics. Due to limitations in bandwidth, the L1 trigger uses coarse information from trigger primitives (TPs) from the calorimeter and muon systems, and the tracker information is only read out if the trigger makes a positive decision. Hence the L1 trigger can be split into the Calorimeter and the Muon Trigger. The Calorimeter Trigger consists of two layers. It is responsible for the reconstruction, calibration and baseline identification of particles which leave energy deposits in the ECAL or HCAL: electrons/photons, hadronically decaying taus and hadronic jets. The hadronic jets are identified as deposits from hadronic interactions located within a narrow geometrical cone. The transverse momenta of the reconstructed jets are used to define the missing transverse energy, which is a measure of the energy imbalance in the detector. It can come from non-interacting, uncharged particles such as neutrinos or possible new particles beyond the Standard Model. As there is no tracker information at this stage, electrons cannot be distinguished from photons, so they are referred to as e/γ objects. The Global Muon Trigger uses the information obtained from the three muon subsystems. Both triggers use information provided by trigger primitives which are formed from basic detector units. Based on the information from the Calorimeter and the Muon Triggers, the Global Trigger then makes the trigger decisions based on a menu of sophisticated algorithms known as the trigger menu.

The central component of the L1 trigger is a set of field programmable gate arrays (FPGAs), which are custom integrated circuits that can be configured after being manufactured. They can perform fast parallel processing and are re-programmable, which enable making changes to the trigger menu to keep it up to date with the evolving physics programme and detector performance. They are implemented in trigger boards which process the data in stages before the Global Trigger makes the final decision using different algorithms. For example, an algorithm can require a specific missing transverse energy threshold for the event.

The L1 trigger menu contains a set of algorithms (referred to as seeds) which are combined through a logical “or”. In order to limit the total rate of events selected by the menu to be under 100 kHz, each seed in the menu has a “prescale” option. It provides the abil-

ity of the trigger to limit the rate of events passing the algorithm by scaling it down by a value. For example, setting a prescale value of 10 leads to one every tenth event which satisfies the conditions imposed by the seed to be passed on to the next stage.

The second level of selection is the HLT [93], which is based on more sophisticated software algorithms and it includes information from the inner tracking system. With the reduced input event rate, it allows for the decision to be made in a window of a few hundred millisecond. This allows for more complex object reconstruction, so the HLT uses a simplified version of the Particle Flow (PF) algorithm [94] to utilise the full detector information. The PF algorithm will be discussed in more detail in Chapter 5. The HLT software is structured in paths which are sequences of algorithmic steps designed to reconstruct physics objects and make selections based on specific physics requirements. Similar to the L1 trigger, the HLT selects events for storage by using a trigger menu, in which the collection of HLT paths is configured. The paths can also be prescaled to reduce the processing time and storage rate.

Chapter 5

Event reconstruction

5.1 Introduction

This chapter discusses the identification and reconstruction of particles and their properties based on the signatures they leave in the detector. Information from all subdetectors is combined using the PF algorithm [94, 95, 96]. It is used to provide a global event description by identifying all stable particles produced during the collision. Due to the fine spatial granularity of the CMS detector, it is well suited for PF reconstruction where the merging of signals from different particles is minimised. This chapter describes the reconstruction of physics objects including muons, electrons, jets and missing transverse energy. The reconstruction of physics objects starts with the reconstruction of tracks, vertices and calorimeter clusters. They are then used in PF as building blocks for physics objects in a global event reconstruction.

5.2 Particle Flow reconstruction

5.2.1 Track reconstruction

With the typical instantaneous luminosities achieved at the LHC, about a thousand charged particles traverse the tracker at each bunch crossing. Therefore it is a challenging environment for maintaining a high track-finding efficiency while ensuring the tracking software runs fast enough at the HLT. It is also important to keep the fraction of “fake” reconstructed tracks small, where “fake” tracks are tracks resulting from a combination of unrelated hits.

The first step of the reconstruction process is called local hit reconstruction, which involves clustering zero-suppressed signals left by particles in pixel and strip modules into hits, and estimating the cluster positions and their uncertainties. The hits can then be used

to estimate the charged particle position and momentum. The hit efficiency, which is defined as the probability to find a cluster in a given silicon sensor that a charged particle has traversed, is on average $> 99\%$ [80].

To achieve high tracking efficiency, an iterative approach is employed in reconstructing tracks. Twelve iterations of a combinatorial track finder algorithm are used for track reconstruction. In each iteration, hits assigned to the tracks in the previous iteration are removed from the hit collection to reduce the combinatorial complexity. It is a good approach for removing easy-to-reconstruct tracks in the first few iterations, leaving the more challenging types of tracks such as soft or very displaced tracks for later stages.

The first four iterations aim to find prompt tracks, which are tracks originating close to the interaction point, while iterations 5-9 find tracks that originate outside the beam spot. The beam spot is a luminous region over which primary vertices from proton-proton collisions are distributed. Iterations 10-12 target tracks of high- p_T jets and muons. Each iteration involves four stages. Firstly, seed generation provides initial track candidates found using a few (2 or 3) hits. A seed gives the initial estimate of the trajectory parameters and their uncertainties. Track finding is then performed using a Kalman filter [97], which extrapolates the seed trajectories along the expected flight path of a charged particle to search for additional hits that can be assigned to the track candidate. The third step involves using the track-fitting module to provide the best possible estimate of the parameters of each trajectory with a Kalman filter and smoother. Finally, track selection sets quality flags and discards tracks that fail certain specified criteria. The main differences between the twelve iterations are the configuration of the seed generation and the final track selection [80].

The tracking efficiency as a function of the track p_T and the simulated track production vertex radius is shown in Fig. 5.1 for simulated $t\bar{t}$ events in Run 2 data-taking conditions. From Fig. 5.1 (right), it can be seen that the later tracking iterations are essential for obtaining good reconstruction efficiency for displaced tracks, which are crucial for searching for long-lived decays.

5.2.2 Vertex reconstruction

The main goal of the vertexing algorithm is to identify the primary vertex (PV) and all tracks associated with it, exclude particles originating from additional collision vertices (pileup), and identify secondary vertices (SVs), which are crucial for long-lived particle searches.

The PV reconstruction consists of three steps. Firstly, tracks consistent with being produced promptly in the primary interaction region are chosen. The selected tracks that

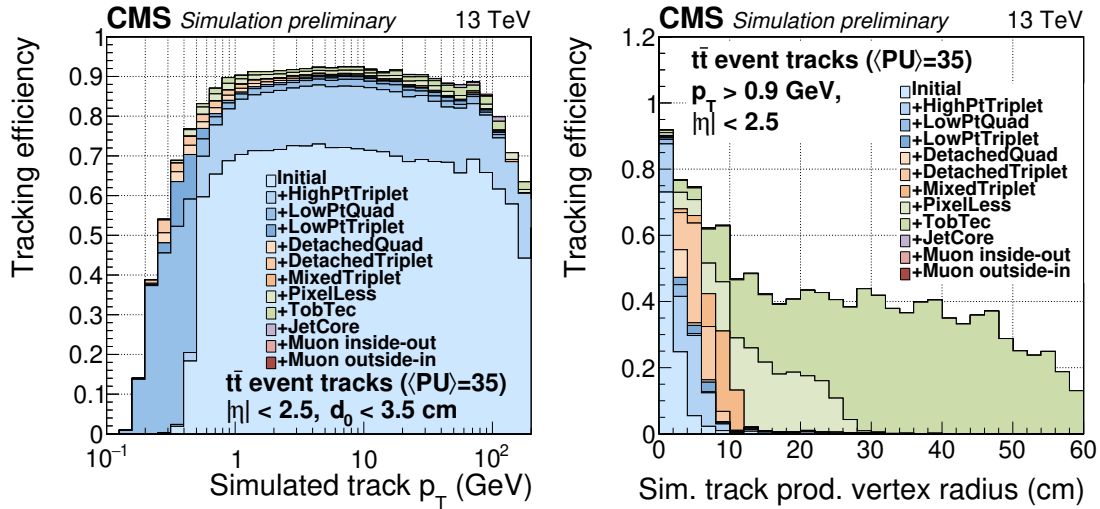


Figure 5.1: Track reconstruction efficiency as a function of simulated track p_T (left) and production vertex radius (right). The efficiencies are shown successively for each of the twelve tracking iterations, which are indicated by different colours [98].

appear to originate from the same interaction vertex are then clustered, based on their z -coordinates at their closest approach to the centre of the beam spot. This allows for the reconstruction of any number of proton-proton interactions in the same LHC bunch crossing. The clustering is performed using a deterministic annealing algorithm [99], which finds the global minimum in a way that is analogous to that of a physical system approaching a state of minimal energy through a series of gradual temperature reductions. Finally, the candidate vertices identified by the deterministic annealing which contain at least two tracks are fitted to determine the position and the corresponding uncertainty of each vertex, as well as the indicators for the success of the fit [100].

The resulting vertex with the highest p_T^2 sum of the associated tracks is chosen to be the leading PV, which is used for physics object identification and reconstruction. All other PVs are treated as pileup interactions. The PV reconstruction efficiency is estimated to be close to 100% when more than two tracks are used to reconstruct the vertex [80].

5.2.3 Cluster reconstruction

The calorimeters are used to reconstruct the energies of incident particles using shower deposits produced in the active material. This is essential as energy clusters left in the calorimeter system are the only means in CMS to reconstruct photons and neutral hadrons, which are uncharged particles. They are identified in the PF algorithm by finding energy clusters that are not associated with the extrapolated trajectory of a charged particle's track.

The clustering algorithm in the ECAL will be discussed as an example. It is used to

sum together energy deposits in adjacent crystals which belong to the same electromagnetic shower. The algorithm starts by grouping together crystals with energies exceeding a predefined threshold, which is generally 2 or 3 times bigger than the electronic noise expected for these crystals. A seed cluster is then identified as the one containing most of the energy deposited in any specific region. The ECAL clusters within a certain geometric area around the seed cluster are then combined into superclusters (SC) in order to include photon conversions and bremsstrahlung losses. This procedure is referred to as “superclustering”. The energy is measured by summing together all the energy deposits recorded within the supercluster’s crystals. The energy collected by the superclustering algorithm is subject to losses for several reasons, for example through lateral and longitudinal shower leakage, in intermodule gaps or dead crystals, and in the tracker. An MVA regression technique is used to derive corrections for the energy of electrons and photons obtained from the superclustering algorithm [101].

5.2.4 Muon reconstruction

Muon identification in the PF algorithm uses information from the tracker and the muon chambers. Tracks are first reconstructed independently in the inner tracker (to form tracker tracks) and in the muon system (to form standalone-muon tracks), and they are then used as input for muon track reconstruction. Standalone-muon tracks are built by collecting information from the muon subdetectors along a muon trajectory using a Kalman-filter technique [97]. The reconstruction starts from seeds made up of groups of segments from the drift tube (DT) or cathode strip chambers (CSC) subdetectors. Each segment is built from the reconstructed hits inside the subdetectors [88].

Tracker muon tracks are built “inside-out” by propagating tracker tracks to the muon system with loose matching to DT or CSC segments. If at least one muon segment matches the extrapolated track, then the tracker track qualifies as a tracker muon track. On the other hand, global muon tracks are built “outside-in” by matching standalone-muon tracks with tracker tracks. The matching is performed by propagating the two tracks onto a common surface and comparing their parameters. A combined fit is performed with the Kalman filter using information from both the tracker track and the standalone-muon track [88].

Due to the high efficiency of the tracker track and muon segment reconstruction, about 99% of the muons produced within the geometrical acceptance of the muon system are reconstructed as a tracker muon track or a global muon track, or very often both. The muons considered in the analysis are required to be either tracker or global muons. Muons reconstructed only as standalone-muon tracks have worse momentum resolution and a higher fraction of cosmic-ray muon contamination than global or tracker muons [88].

A set of variables is used to define various types of muons which are used in analyses.

Some variables are based on muon reconstruction, while others use inputs from outside the reconstructed muon track such as compatibility with the primary vertex. Examples of the main types of muons include loose muon identification (ID) and tight muon ID. A loose muon is a muon selected by the PF algorithm that is also either a tracker or a global muon. A tight muon is a loose muon with a tracker track that uses hits from at least six layers of the inner tracker, including at least one pixel hit. Contrary to the loose ID, the muon must be reconstructed as both a tracker and a global muon. The tracker muon must have matching segments in at least two of the muon stations, while the global muon fit must have $\chi^2/\text{dof} < 10$ and include at least one hit from the muon system. Furthermore, a tight muon must be compatible with the primary vertex, have a transverse impact parameter $|d_{xy}| < 0.2 \text{ cm}$ and a longitudinal impact parameter $|d_z| < 0.5 \text{ cm}$ [88]. The impact parameters are defined with respect to the primary vertex.

The loose ID is used for the analysis as it has no requirements on muon isolation or displacement, which is suitable for searching for displaced decays with high multiplicity. Tight ID could be suitable if the requirements on the impact parameters were removed, but it would require custom changes and calibrations so it is not used. The efficiency of muon reconstruction and loose identification are shown in Fig. 5.2 for muons with $p_T > 20 \text{ GeV}$, measured using the tag-and-probe technique [88]. It can be seen that the efficiency is close to 100%.

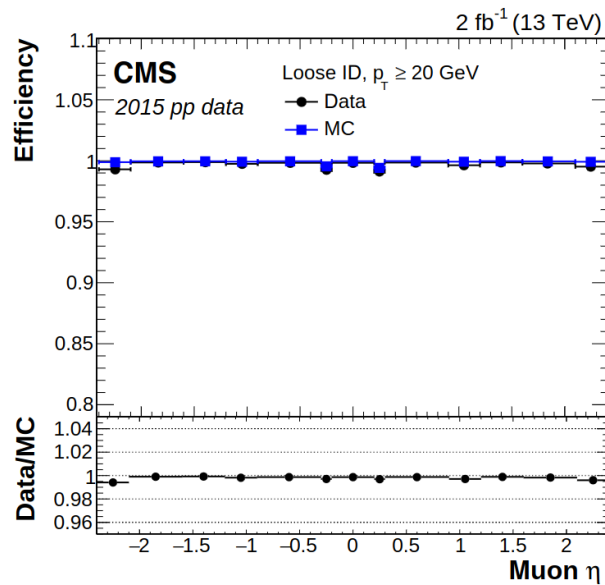


Figure 5.2: Muon reconstruction and identification efficiency in 2015 data and simulation for loose muons with $p_T > 20 \text{ GeV}$ [88].

5.2.5 Electron reconstruction

Electron candidates are required to satisfy specific identification criteria, which can either be cut-based selections (a sequential set of requirements) or based on multivariate techniques.

For the cut-based selections, non-prompt electrons are rejected using different variables, for example the ratio of hadronic to electromagnetic energy H/E , and the second moment of the log-weighted distribution of crystal energies in η , denoted as $\sigma_{i\eta i\eta}$. Additional variables related to the track properties are also used. One of them is $|1/E - 1/p|$, where E is the supercluster energy and p is the track momentum at the point of closest approach to the vertex. Angular variables are also used, which are $|\Delta\eta_{\text{in}}^{\text{seed}}| = |\eta_{\text{seed}} - \eta_{\text{track}}|$ (where η_{seed} is the η position of the seed cluster and η_{track} is the track η extrapolated from the innermost track position) and $|\Delta\phi_{\text{in}}| = |\phi_{\text{SC}} - \phi_{\text{track}}|$ (where ϕ_{SC} is the supercluster energy-weighted position in ϕ). An important source of background to prompt electrons stems from secondary electrons produced in conversions of photons in the tracker material. The minimum number of missing hits in the tracker is used to reject this background. Furthermore, a pixel conversion veto is applied. Details of all cut-based electron identification requirements can be found in Ref [101].

In terms of multivariate techniques, a Boosted Decision Tree (BDT) is used to combine several variables in order to further improve the performance of the electron identification. The set of variables is extended compared to the cut-based selections. To maximize the flexibility at analysis level, the electron identification BDTs are trained with and without including the isolation variables. Although the BDT method provides better background rejection for a given signal efficiency compared to the cut-based approach, it is easier to change a specific selection for the latter. The electron identification efficiency for the cut-based approach and the BDT-based approach are shown in Fig. 5.3, which are measured using the tag-and-probe technique [101].

5.2.6 Jet reconstruction

Collimated jets of particles are produced by the hadronisation of partons (quarks and gluons) coming from deep inelastic proton-proton collisions. In CMS, PF candidates are clustered into jets using the anti- k_t [102] algorithm. It satisfies infrared safety (i.e. not being sensitive to the addition of soft particles) and collinear safety (i.e. not being sensitive to the collinear splitting of particles), which lead to theoretically tractable jet properties.

In the jet clustering algorithm, the distance measures d_{ij} and d_{iB} are introduced, which represent the distance between particles i and j , and the distance between particle i and

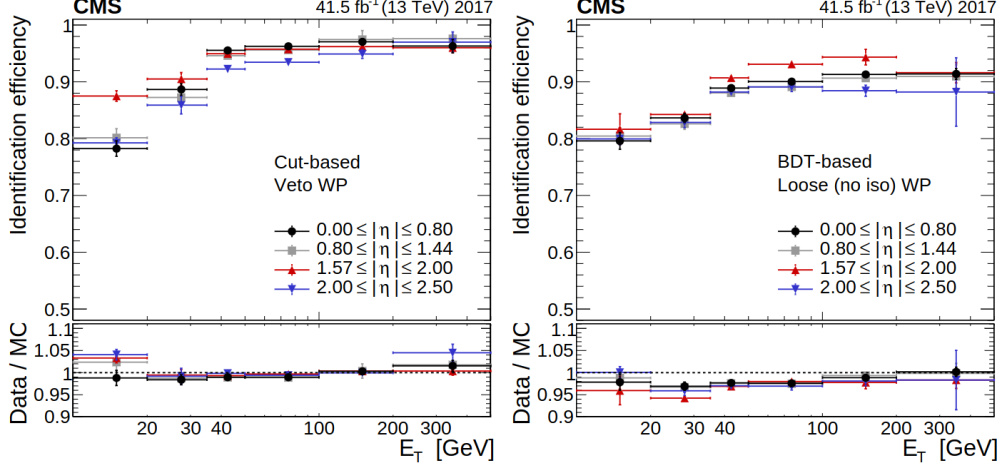


Figure 5.3: Electron identification efficiency measured in data (upper panels) and data-to-simulation efficiency ratio (lower panels) as a function of the electron E_T for a cut-based identification working point (left) and a BDT-based working point (right) [101].

the beam respectively. They are defined in the anti- k_t algorithm as:

$$d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \quad (5.1)$$

$$d_{i,B} = p_{T,i}^{-2}, \quad (5.2)$$

where $p_{T,i}$ is the transverse momentum of particle i , ΔR_{ij} is the ΔR between particles i and j defined as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, and R is a radius parameter that controls the size of the jet.

The clustering is performed by identifying the smallest of the distances d_{ij} and $d_{i,B}$. If the smallest distance is d_{ij} , particles i and j are merged into a single entity. If the smallest distance is $d_{i,B}$, then i is identified as a jet and it is removed from the list of entities. The distances are recalculated and the procedure repeated until there are no entities left.

A common choice of the radius parameter in the algorithm is $R = 0.4$, which leads to “AK4 jets”. These jets are used for a study of the input variables of the BDT which is discussed in Section 7.3.2. Charged jet constituents that do not belong to the primary vertex are removed from the jet to reduce pileup. This is performed with a method known as charged hadron subtraction [103]. In order to match the energy of the reconstructed jet to that of the parton that produced it, the jet four-momentum is calibrated by applying jet energy corrections, which are derived from simulation and adjusted with data [104].

Additional jet identification criteria are applied to create analysis level objects. The exact requirements are adjusted slightly for each year of data taking. Typically, jets are required to contain at least two constituents. The neutral hadron fraction and the neutral electromagnetic energy fraction are usually required to be both less than 90%. Moreover, jets are

typically required to be within the tracker acceptance with at least one constituent being a charged hadron, and the charged electromagnetic energy fraction is required to be less than 99%.

5.2.7 Missing energy reconstruction

As the system of colliding partons has negligible initial transverse momentum, the transverse momenta of all final state particles must add up to zero within the detector resolution. Any deviation from this is known as missing transverse energy (MET), denoted by \vec{p}_T^{miss} . Aside from neutrinos, it could come from weakly interacting particles beyond the Standard Model, so it is an important quantity. It is defined as the negative vector sum of the transverse momenta of all PF candidates in an event:

$$\vec{p}_T^{\text{miss}} = -\sum_i \vec{p}_T^i \quad (5.3)$$

It is also convenient to define the scalar quantity $p_T^{\text{miss}} = |\vec{p}_T^{\text{miss}}|$. The raw MET is biased by effects like minimum calorimeter energy thresholds, non-linearity of calorimeter responses and inefficiencies in track reconstruction [105]. These are accounted for by including jet energy corrections (JECs) in the calculation:

$$\vec{p}_T^{\text{miss,corr}} = \vec{p}_T^{\text{miss}} - \sum_{\text{jet}} (\vec{p}_{T,\text{jet}}^{\text{corr}} - \vec{p}_{T,\text{jet}}), \quad (5.4)$$

where jets with $p_T > 15 \text{ GeV}$ that have less than 90% of their energy deposited in the ECAL are used. This requirement ensures that electrons and photons that are misidentified as jets are not considered in the correction. Figure 5.4 shows the PF MET distributions in $Z \rightarrow \mu^+ \mu^-$ and $Z \rightarrow e^+ e^-$ events, including uncertainties from JECs.

Further corrections are applied to account for pileup effects, which are performed using the PUPPI algorithm [103]. It aims to use information related to local particle distribution, pileup properties and tracking information to mitigate the effect of pileup. The algorithm operates at the particle candidate level, before any clustering is performed. It calculates a weight in the range of 0 to 1 for each particle using information about the surrounding particles, where a value of 1 is assigned to particles considered to originate from the primary vertex. These weights are used to rescale the particle four-momenta to correct for pileup at the particle level so as to reduce the contribution of pileup, such that $\vec{p}_T^{\text{miss}} = -\sum_i w_i \cdot \vec{p}_T^i$.

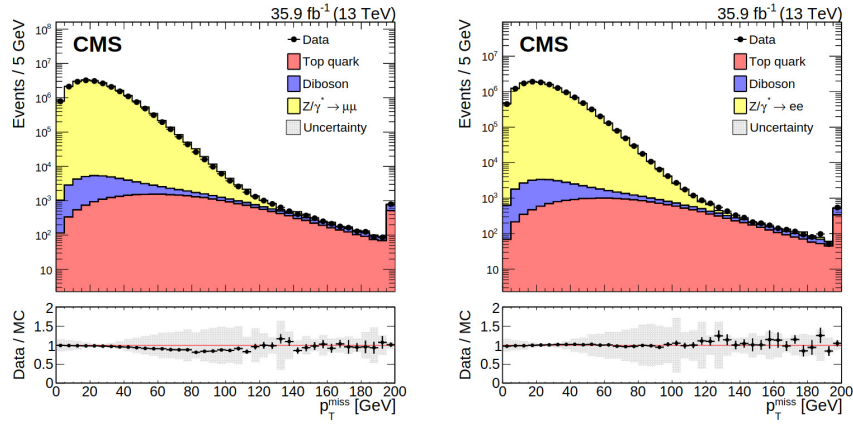


Figure 5.4: Distributions of p_T^{miss} in $Z \rightarrow \mu^+ \mu^-$ (left) and $Z \rightarrow e^+ e^-$ (right) in data and simulation. The last bin includes all events with $p_T^{\text{miss}} > 195$ GeV [105].

Chapter 6

Neural Network Vertexing in the Phase-2 upgrade of the Level-1 Trigger

6.1 Introduction

The upgrade to the High-Luminosity Large Hadron Collider (HL-LHC) [106] will provide an unprecedented level of luminosity, resulting in an integrated luminosity of 3000 fb^{-1} in the baseline configuration, which brings immense opportunities for physics analyses. With the large data samples collected, a wide range of analyses will become possible. Extended searches for new physics beyond the Standard Model and unprecedented high-precision measurements of the Standard Model will be performed, which include significant improvements in the measurements in the Higgs sector. Ultimately the HL-LHC can reach a peak instantaneous luminosity of $7.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, and the average number of proton-proton collisions per bunch crossing (pileup) will increase to around 200. In the ultimate configuration, the integrated luminosity can reach 4000 fb^{-1} at the end of the HL-LHC lifetime.

The increased collision rate and higher level of pileup lead to high particle multiplicity and an intense radiation environment for the detector. Upgrades of different parts of the CMS detector are going to be performed to fully exploit the potential of the HL-LHC [107]. In particular, the CMS trigger system will be upgraded to maintain high performance in the challenging running conditions [108], and also to significantly improve the selection of any possible new physics that could lead to unconventional experimental signatures.

One of the most important upgrades is the inclusion of state-of-the-art techniques used in offline reconstruction and analyses at the Level-1 (L1) trigger, for example event reconstruction based on particle-flow techniques. The Phase-2 upgrade of the trigger and the data acquisition system will remain a two-level system. In terms of the hardware imple-

mentation, Field Programmable Gate Arrays (FPGAs) coupled with high-speed optical links will be used to retrieve the detector data and provide the interconnections between the systems. The L1 maximum rate will be increased to 750 kHz, and the total latency will be increased to $12.5 \mu\text{s}$ to allow the inclusion of information from the tracker [109] and the high-granularity calorimeter [110] for the first time. The longer latency will enable higher-level object reconstruction and identification, and also the evaluation of complex event quantities to optimise physics selections. Therefore, the implementation of sophisticated algorithms such as particle-flow reconstruction and machine learning techniques becomes possible at the L1 trigger, which will bring great improvement to the performance of object reconstruction in the challenging environment.

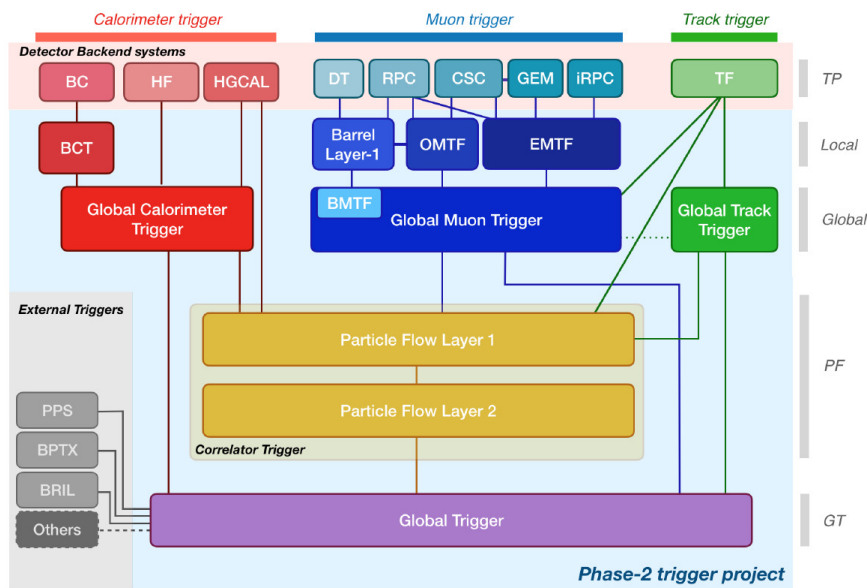


Figure 6.1: The Phase-2 upgrade design for the CMS Level-1 trigger. The trigger uses information from the calorimeters, the muon spectrometers and the track finder. The correlator trigger makes up of two layers that are dedicated to particle-flow reconstruction. All reconstructed objects are sent to the global trigger, which makes the final Level-1 trigger decision [108].

The key feature of the Phase-2 upgrade of the L1 trigger system is the introduction of a correlator layer as shown in Fig. 6.1, which uses information of multiple sub-detectors and implements sophisticated algorithms to produce higher-level trigger objects. This results in improved selectivity of the L1 trigger that approaches the performance of the High-Level Trigger. Four independent data processing paths are implemented for optimum flexibility, which are tracking, calorimetry, muon systems and particle-flow techniques.

L1 tracks delivered by the upgraded Outer Tracker (OT) will become available at the L1 trigger. A new type of module is used for the OT in Phase-2 to reduce the data rate. It consists of two closely spaced layers of silicon with a tunable window which allows for an on-module cut for charged particles with $p_T > 2 \text{ GeV}$. When a charged particle traverses

the CMS detector, it curves in the magnetic field and the radius of curvature is inversely proportional to the particle's p_T . As shown in Fig. 6.2, a high p_T particle passing through the two silicon layers that falls within the window shown in green (the leftmost track) is registered as a “stub” for L1 reconstruction. Low p_T , curved particle trajectories that fall outside the window (the rightmost track) are not used for L1 reconstruction. The hits in the detector layers are stored regardless of their p_T so that if the event passes the L1 trigger, the full p_T spectrum of tracks can be readout and reconstructed in the HLT and offline [109].

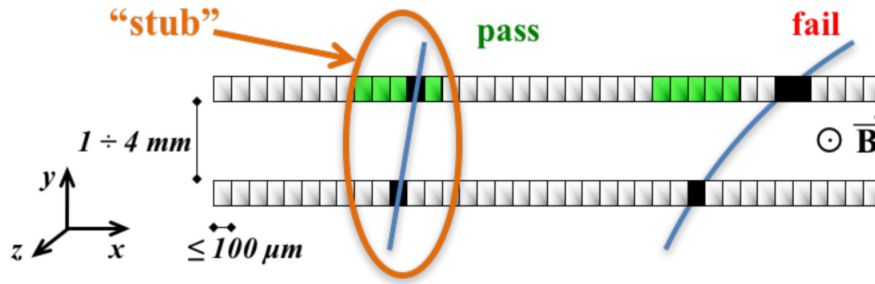


Figure 6.2: Illustration of the upgraded outer tracker p_T module concept. Correlation of signals in closely-spaced sensors allows rejection of low- p_T particles. The selection window shown in green is used to define an accepted stub [109].

The L1 tracks are reconstructed in the track finder processors as part of the detector backend. The track parameters and quality flags from the track reconstruction are supplied to the trigger system to perform precise vertex reconstruction and matching with calorimeter and muon objects. This maximises the trigger efficiency and keeps the trigger rate down. A global track trigger (GTT) will be included to perform reconstruction of primary vertices, displaced vertices, as well as tracker-only based objects such as jets and missing transverse momentum.

6.2 Baseline approach for vertex reconstruction

The identification and reconstruction of the primary vertex from tracks are vital for the mitigation of pileup in the upgraded detector, as well as the efficient use of resources. The goal is to identify the hard scatter vertex and its associated tracks or particle-flow candidates. Here the primary vertex is taken to be the reconstructed vertex with the largest sum of track p_T (while in offline reconstruction it is defined as the vertex with the highest sum of p_T^2). By selecting the tracks associated to the primary vertex, the downstream algorithms can be targeted towards the interactions coming from the hard scatter, hence mitigating pileup and reducing resource usage.

The baseline approach is a histogram-based algorithm called "FastHisto". The algorithm creates a histogram of track z_0 where each track is weighted by its p_T . The track z_0 is

calculated from the Kalman Filter (KF) z_T subtracting a factor based on the $\cot(\theta)$ of the track. The z_T is the z -coordinate of the track when it crosses a cylinder of radius T where the density of stubs inside and outside the cylinder is equal. To determine the position of the vertex, a three-bin window is passed across the histogram to find the three consecutive bins with the maximum scalar sum p_T . The centre of the three bins is returned as the position of the primary vertex. An illustration of the algorithm is shown in Fig. 6.3. There are several shortcomings of the FastHisto algorithm. For example, there is no handle on fake tracks so high p_T fakes together with pileup tracks can enter the histogram and result in a fake vertex. Also, it does not correct for the degradation in z_0 resolution with increasing track $|\eta|$, which worsens the resolution of the reconstructed primary vertex.

After the vertex position is obtained, track-to-vertex association is performed in the baseline approach using the distance between the primary vertex and the track z_0 , denoted by Δz_0 . The resolution of the track z_0 position is highly dependent on $|\eta|$, and it increases with $|\eta|$ approximately as $\cosh(|\eta|)$ due to the change of the resolution of the tracks based on their orientation along the barrel shape of the tracker. Therefore, instead of using a single z_0 threshold to associate tracks to the vertex, the thresholds are binned in $|\eta|$ to allow a larger Δz_0 threshold for tracks with higher $|\eta|$. This is found to be a reasonably effective way to perform track-to-vertex association, with reasonable performance in associating genuine primary vertex tracks instead of fake or pile-up tracks. However, it is expected that the performance can be improved by using other track features to better discriminate between pileup and primary vertex tracks, for example by using the quality of the tracks.

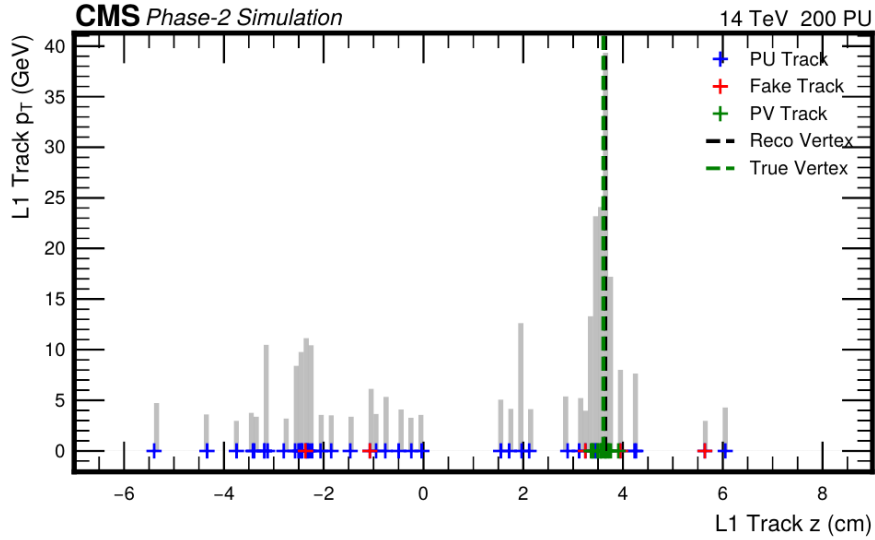


Figure 6.3: Vertex reconstruction with the FastHisto algorithm, which uses a p_T weighted histogram of the track z position. The coloured crosses represent the actual position of each track for different types of tracks, which are pileup tracks, fake tracks and primary vertex tracks. The grey markers show the scalar sum of the track p_T in each bin. The black dashed line indicates the vertex position found by the FastHisto algorithm, and the green dashed line shows the position of the true vertex [108].

6.3 End-to-End Neural Network Vertexing

6.3.1 Network architecture and learning

To improve the performance of the vertex finding, more complex weightings can be applied to the tracks in the histogram in the baseline approach. For example, cuts on the track quality MVA score can remove fake tracks. The track quality MVA is the output from a Boosted Decision Tree that is trained using track features such as track χ^2 and the number of stubs to discriminate between real and fake tracks. Also, a weighting such as $\frac{1}{\eta^2}$ can take into account the degradation in z_0 resolution in higher $|\eta|$. To optimise the use of these features in calculating the weightings, a dense neural network is employed. Moreover, the calculation of the vertex position from the weighted histogram can be improved. A 1D convolutional neural network is used instead of the three-bin window in the baseline FastHisto approach to allow more complex pattern recognition. Finally, aside from the vertex finding, the track-to-vertex association can also benefit from the use of more complex track features such as the track quality MVA, in addition to the Δz_0 between the track and the vertex. As the track-to-vertex association can be improved by the correct reconstruction of the primary vertex, and the association of tracks to the vertex is ultimately the most important output to downstream algorithms, an end-to-end approach is implemented to learn simultaneously the event-level primary vertex regression and the track-level track-to-vertex association.

The end-to-end network consists of three parts to perform primary vertex regression and track-to-vertex association. The weight function dense neural network (DNN) learns track weights using track features, which are the track p_T , η and track quality MVA score. The learned track weights are passed to a pattern recognition convolutional neural network (CNN). An ArgMax, which returns the z position of the maximum bin in the histogram, is used to determine the vertex z position. Using the track input features and the distance of the track from the primary vertex, an association dense neural network is implemented to perform track-to-vertex association. The network architecture is shown in Fig. 6.4 [111].

The network is trained with two loss functions for the vertex regression and the track-to-vertex association, which are equally weighted. A Huber loss function is attributed to the vertex regression. The loss function is defined as

$$L_\delta(z_{\text{true}} - z_{\text{predicted}}) = \begin{cases} \frac{1}{2} (z_{\text{true}} - z_{\text{predicted}})^2 & \text{if } |(z_{\text{true}} - z_{\text{predicted}})| \leq \delta \\ \delta (|(z_{\text{true}} - z_{\text{predicted}})| - \frac{1}{2}\delta) & \text{otherwise,} \end{cases} \quad (6.1)$$

where z_{true} is the z position of the true vertex, $z_{\text{predicted}}$ is the z position of the predicted

vertex, and δ is a tunable parameter called the Huber delta that controls the overall shape of the loss function. The loss is quadratic when the predicted z value is within δ of the true value, and it is linear when the difference between the predicted and the true z is larger than δ . By tuning the Huber δ to a higher value, the tails of the residuals can be improved as more residuals will be controlled by the quadratic loss function.

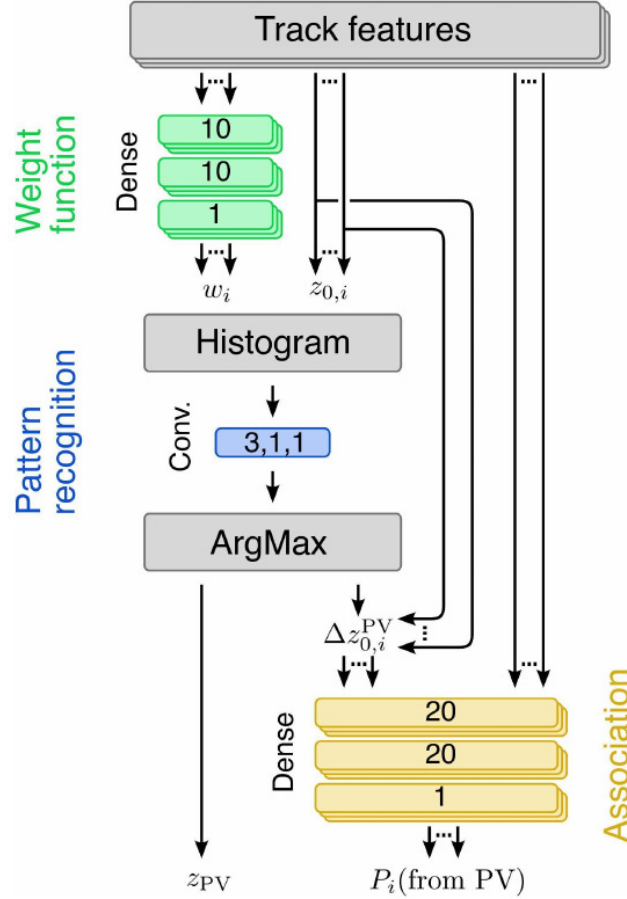


Figure 6.4: Network architecture for the End-to-End Neural Network Vertexing, which learns simultaneously primary vertex regression and track-to-vertex association. Track features are first fed into a dense neural network that learns a weight for each track. The weights are used to fill a histogram in z_0 , which is then passed through a convolutional neural network to perform peak-finding. The z position of the primary vertex is returned with an ArgMax, which finds the maximum bin of the histogram. The track’s distance from the primary vertex $\Delta z_{0,i}^{PV}$ and the input track features are then passed to a second dense neural network to learn the probability of the track associated to the regressed primary vertex [111].

A binary cross entropy loss function is used for the track-to-vertex association. For an event with N tracks, each with a predicated association probability of p_i and an actual class label of y_i , the loss function is defined as

$$L(\mathbf{y}) = \frac{1}{N} \sum_{i=1}^N -(y_i \log(p_i) + (1 - y_i) \log(1 - p_i)). \quad (6.2)$$

Apart from the loss functions used for the vertex regression and track-to-vertex association, auxiliary losses are also included to improve the performance of the network. The lasso regularisation (also known as the L1 regularisation) allows the reduction of weights down to zero by adding a loss term that is proportional to the sum of the magnitudes of all weights [112]. This makes the subsequent pruning of the network more effective. Moreover, the ridge regularisation (also known as L2 regularisation) is applied to reduce the magnitudes of the weights in the network [113], which makes subsequent network quantisation more effective. It adds a penalty term that is proportional to the sum of the squared magnitudes of the weights, and so it penalises weights with higher values more than those with smaller values.

A final loss function is added to control the magnitude of the learned track weights which are used to fill the histogram. The loss function finds the top 10% of the learned track weights and returns a loss proportional to the mean square of these weights, hence penalising learned track weights with high magnitude. By constraining the magnitude of learned track weights, the neural network training becomes more reproducible and it is easier to perform quantisation of the network.

As the end-to-end neural network is trained to simultaneously learn the vertex regression and track-to-vertex association, every part of the network needs to be differentiable for the back propagation. Therefore, a custom TensorFlow layer is used for the histogram when performing vertex finding. Each bin of the histogram h_i is filled with track z_0 weighted by weight w , which is a learned function of the track p_T, η and MVA score. There are 256 bins in the histogram of z_0 . In other words, each histogram bin is the weighted sum of all tracks that fall into that z_0 bin, which can be expressed as

$$h_i = \sum_{j=0}^{\text{tracks}} \delta(j \in \text{bin } i) w(p_{Tj}, \eta_j, \text{MVA}_j). \quad (6.3)$$

The resulting gradients of the histogram are then

$$\frac{\partial h_i}{\partial z_0} = 0 \quad \text{and} \quad \frac{\partial h_i}{\partial w} = \sum_{j=0}^{\text{tracks}} \delta(j \in \text{bin } i). \quad (6.4)$$

Since the histogram is differentiable, it can be used in both forward and backward passes of the neural network training. The vertex position is obtained by an ArgMax on the histogram filled with the learned track weights. As the standard ArgMax is non-differentiable, a soft ArgMax is used to enable back propagation. It takes an input vector \mathbf{x} with N elements, and is implemented using several standard TensorFlow layers and so it is differentiable. A temperature-rescaled SoftMax is first applied on the input vector to return the

output

$$\frac{e^{x_i/T}}{\sum_{j=0}^N e^{x_j/T}} . \quad (6.5)$$

The output from the SoftMax can be interpreted as a probability distribution of the track weights (with a value between 0 and 1). After that, the output vector is multiplied by the bin index in the histogram, i.e. a single linear layer of weights between 0 and 255 is implemented. The multiplication returns 0 where the SoftMax output is 0, and returns the bin index where the SoftMax output is 1.

Finally, the resulting vector is summed over to give the bin index for which the SoftMax output of 1 is located:

$$\sum_{i=0}^N i \frac{e^{x_i/T}}{\sum_{j=0}^N e^{x_j/T}} \quad (6.6)$$

The equations for the ArgMax are implemented with TensorFlow operations, and it is usable in both forward and backward passes of the training.

In order to avoid overtraining, the training and validation errors are monitored during the neural network training. The training loss and validation mean square error (MSE) in z_0 are shown in Fig. 6.5 for an example training. There are a few fluctuations in the validation MSE, but no obvious overtraining is observed.

6.3.2 Implementation

As the end-to-end neural network is implemented in firmware, the resource usage has to meet the budget of the trigger. Firstly, quantisation is applied to the network layer-by-layer using the QKeras package [114] and the spread of the weights are inspected in every layer. Quantisation is the technique of storing the network parameters in lower numerical precision in order to reduce the computational time and energy [115]. It is noted that an automatic quantisation cannot be implemented because of the custom TensorFlow layers included in the network.

After quantisation, pruning is also applied to further reduce the resource usage. A custom cycle of pruning is implemented due to the presence of non-standard TensorFlow layers. After a cycle of training and the quantisation of the network weights, all the weights close to zero are chosen. They fill a mask that set these weights to zero. The mask is then used together with the previously trained network in a new training cycle where the network is trained with the pruning applied. The use of pruning cycles is more stable than removing all small weights at the end of the training. It also allows more weights to be removed while maintaining performance. Eight pruning cycles are used in total, targeting only the weight and association layers, since the pattern CNN is small and pruning will likely affect its performance.

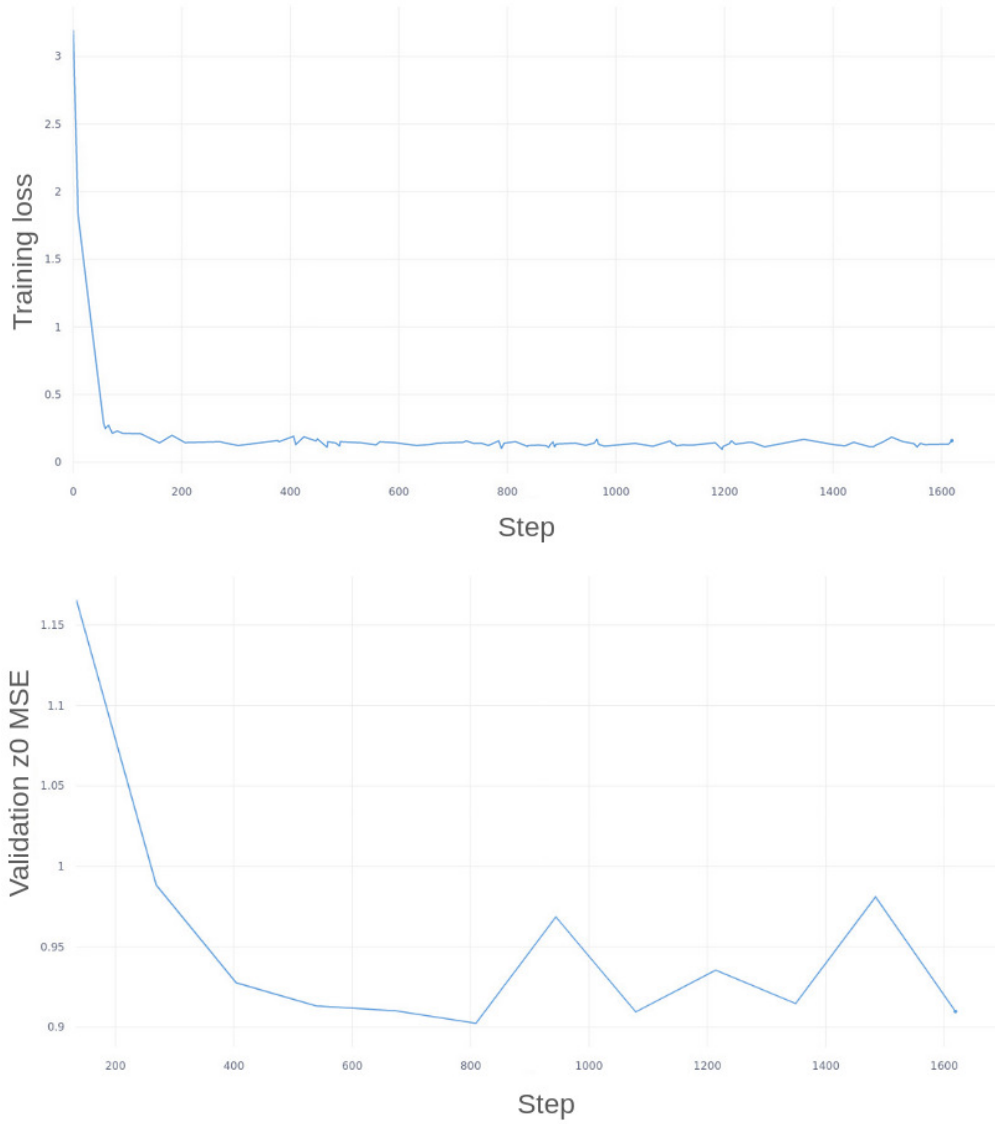


Figure 6.5: Training loss and validation z_0 mean square error in an example training of the neural network.

The end-to-end neural network vertexing is implemented in firmware by inserting the weight DNN, the pattern CNN, and the association DNN within the existing baseline firmware. The weight network and the association network are implemented with 18 copies of each network to match the 18 input track streams. All parts of the end-to-end neural network are implemented in 3 time multiplexed slices in the GTT, resulting in a total of 54 copies of the weight network, 3 copies of the pattern network and 54 copies of the association network implemented on an field-programmable gate array.

6.3.3 Performance

The end-to-end neural network vertexing outperforms the baseline FastHisto approach in both the vertex position regression and track-to-vertex association.

The studies are performed with a sample of $t\bar{t}$ events, first using old KF tracks. The performance of the network trained with new KF tracks is also studied. The new KF is a bit accurate emulation of the KF, which has worse z_0 resolution than the simulated KF due to missing higher order corrections. It also does not apply any internal χ^2 cuts on the tracks, which leads to a higher fake rate. The new KF corresponds closely to the currently available firmware, while the old KF could be regarded as a simulation of a possible future implementation with potential improvements.

The vertex z_0 residuals are shown for the neural networks and the baseline approach using the old KF tracks in Fig. 6.6. An improvement of 34% is achieved in the root mean square (RMS) of the residuals in the neural network vertexing compared to the baseline FastHisto algorithm. The improvement comes from the tails of the distribution, as seen in the left plot of Fig. 6.6. The performance is similar in the core of the residuals, as shown in the right plot of Fig. 6.6. The improvement in vertex reconstruction can also be seen in Fig. 6.7, which shows the z_0 of the reconstructed vertex against the z_0 of the true vertex for the baseline algorithm and the quantised neural network respectively. There is a smaller spread of vertices around the diagonal for the neural network, which shows a higher accuracy in vertex reconstruction. The improvement in the tails is mainly due to the additional track quality information used in the neural network, reducing the misreconstruction of vertices from high p_T fake tracks which is more common in the baseline FastHisto algorithm.

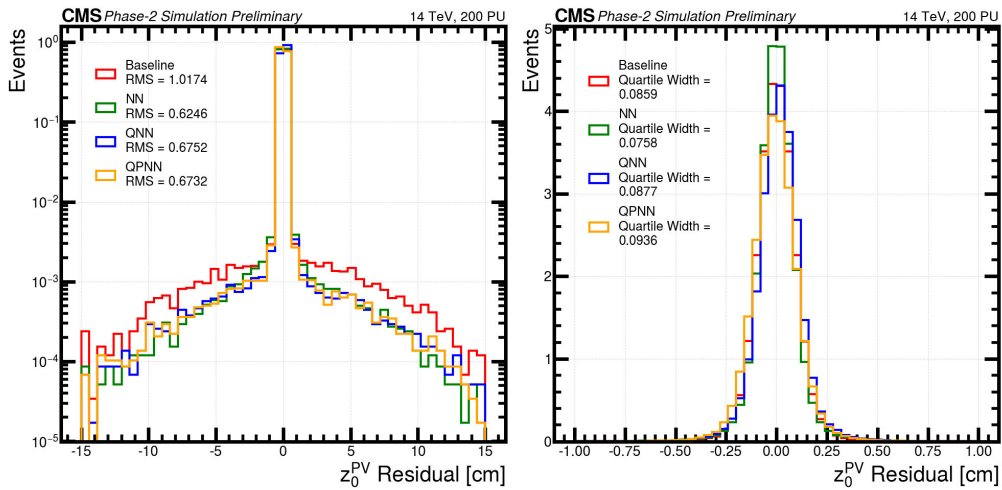


Figure 6.6: z_0 residuals of neural network vertexing trained with the old Kalman Filter tracks for the floating point, quantised (QNN) and quantised pruned (QPNN) neural networks compared to the baseline FastHisto approach.

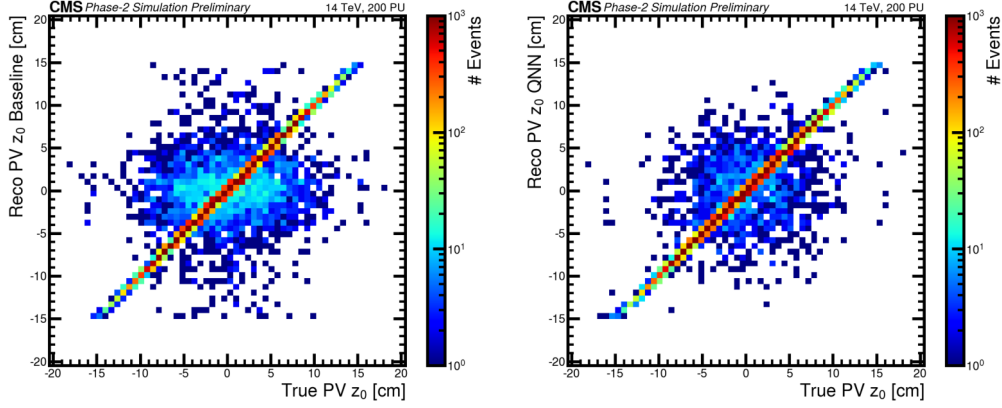


Figure 6.7: Two-dimensional plots showing the z_0 of the reconstructed primary vertex using old Kalman Filter tracks against the z_0 of the true primary vertex for the baseline FastHisto algorithm on the left, and for the quantised neural network on the right. The diagonal strip shows vertices that are correctly reconstructed, while the distribution around the diagonal strip shows vertices which are misreconstructed. The smaller spread of vertices around the diagonal shows a better performance in vertex reconstruction for the neural network compared to the baseline.

The performance in vertex regression is also studied with the new KF tracks, as the new KF matches closely with the currently available firmware. The z_0 residuals of the neural network vertexing and the baseline FastHisto algorithm are shown in Fig. 6.8. There is a big reduction of the residuals especially in the tails for the neural network vertexing compared to FastHisto, with an improvement of 59% in the RMS of the residuals. It is noted that the FastHisto algorithm is not yet optimised for the new KF tracks, while the neural network has the advantage of adapting to the new tracks while it is trained to maintain good performance. There is also an improvement of about 9% in the core of the residuals for the neural network compared to the baseline FastHisto. The reconstructed z_0 against the true z_0 of the primary vertex is shown in Fig. 6.9. There is significant improvement in the reconstruction accuracy, as seen in the more concentrated distribution of vertices along the diagonal for the neural network compared to the baseline FastHisto algorithm. It is also noticed that the baseline approach has a slight bias towards reconstructing primary vertices at z_0 closer to 0, as seen in the green oval shape at the centre of the distribution. The oval shape is not seen in the distribution for the neural network approach, which shows that it gives less biased predictions.

Comparing the z_0 residuals changing from the old KF to the new KF, an increase in the residuals is seen for both the baseline FastHisto algorithm and the neural network. This is expected as the new KF tracks have worse z_0 resolution than the old KF tracks. The core of the residuals increase by 54% for the FastHisto algorithm while the increase is 29% for the neural network when changing from old KF tracks to new KF tracks. The use of track quality MVA as input to the neural network helps mitigate the increased fake rate in the

new KF tracks, leading to better performance compared to the FastHisto. This shows that the neural network is more robust to the change of the underlying track features.

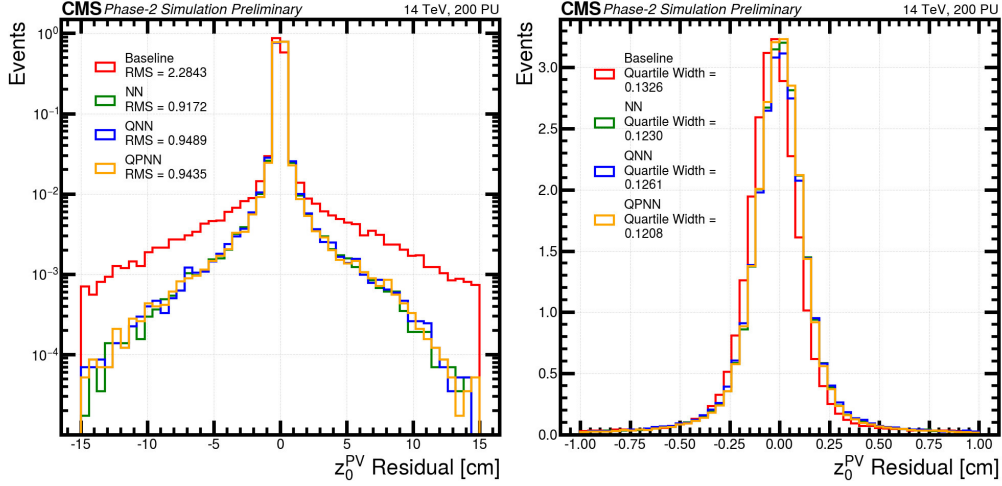


Figure 6.8: z_0 residuals of neural network vertexing trained with the new Kalman Filter tracks for the floating point, quantised (QNN) and quantised pruned (QPNN) neural networks compared to the baseline FastHisto approach.

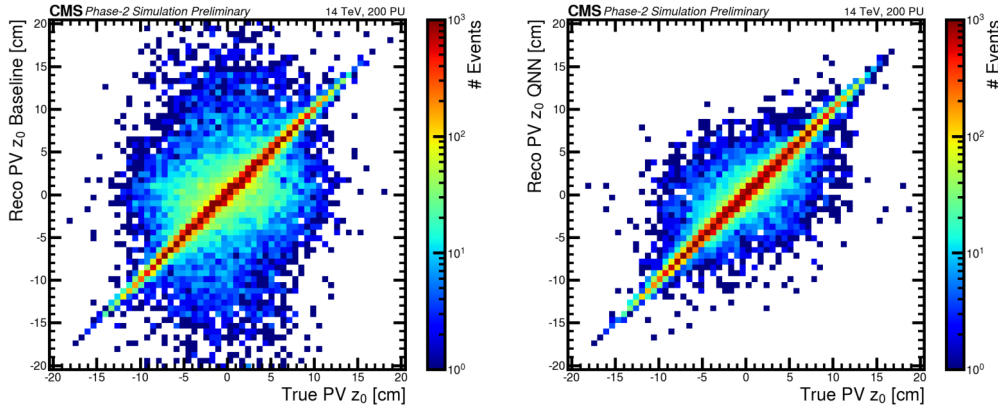


Figure 6.9: Two-dimensional plots showing the z_0 of the reconstructed primary vertex using new Kalman Filter tracks against the z_0 of the true primary vertex for the baseline FastHisto algorithm on the left, and for the quantised neural network on the right. The diagonal strip shows vertices that are correctly reconstructed, while the distribution around the diagonal strip shows vertices which are misreconstructed. The smaller spread of vertices around the diagonal shows a better performance in vertex reconstruction for the neural network compared to the baseline.

The neural network also outperforms the baseline FastHisto in performing track-to-vertex association. The performance is shown in Fig. 6.10 and Fig. 6.11 for the old and new KF tracks respectively. For the same purity, the neural networks have higher efficiencies in associating tracks to the reconstructed vertex compared to the baseline, and vice versa. For the same track to vertex association false positive rate, the neural networks have higher true positive rates. The improvement in the ROC AUC is about 6% in the neural

network compared to the baseline with the old Kalman Filter tracks. The ROC AUC is the area under curve of the receiver operating characteristic curve, which is a plot of the true positive rate against the false positive rate at each neural network threshold score. It is a metric of the performance of the neural network classification. There is a more significant improvement in the ROC AUC of the neural network over the baseline with the new Kalman Filter tracks, which is about 22%. When comparing the ROC AUC of the quantised neural network and quantised pruned neural network to that of the neural network, similar performance is observed. As in the case of vertex regression, the neural network is more robust to the change from the old KF tracks to the new KF tracks when performing track-to-vertex association. This can be seen from the smaller change in the performance in the neural network compared to the baseline with the change of the tracks.

The resource usage of the neural network implemented on a Xilinx VU13P running at 360 MHz for one super logic region (SLR) is shown in Table 6.1. A Vivado implementation is used to estimate the resource usage, and a simulation is used to estimate the latency and initiation interval. Post-training quantisation is applied to reduce usage by tuning the quantisation to the size of which the network weights have converged to. The resource usage is shown for the quantised pruned weight network, the quantised pattern network and the quantised pruned association network. Taking into account the number of track streams and time multiplexes when implemented in the Global Track Trigger, the total resource usage for the networks on a VU13P is estimated to be 28.9% in Look-up tables (LUT), 20.5% in Digital signal processor (DSP) slices and 11.4% in Flip flops (FF).

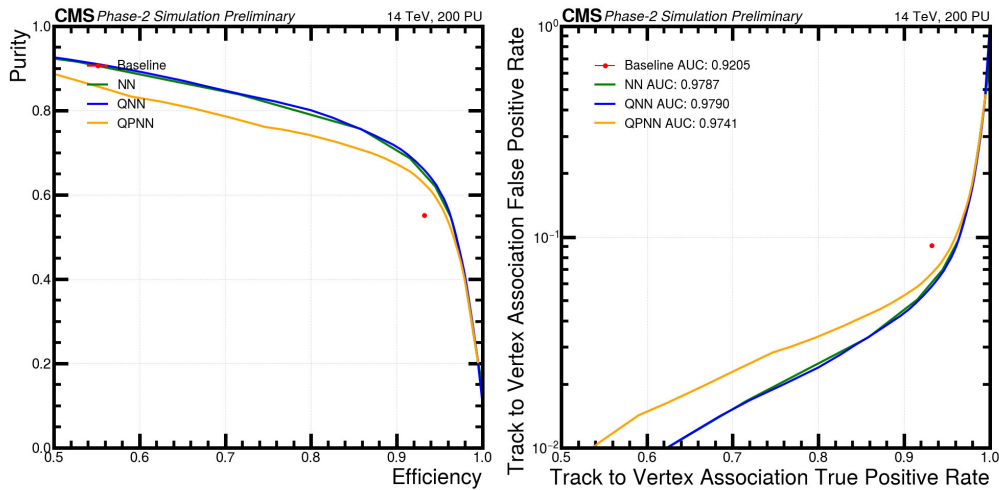


Figure 6.10: Purity against efficiency (left) and receiver operating characteristic (ROC) curves (right) for the baseline cut-based approach and the neural network track-to-vertex association with the old Kalman Filter tracks. The baseline is shown in as a red dot as it represents a single cut, as opposed to varying thresholds for the neural networks. In the ROC plot, the ROC AUC is the area under curve for the neural network approaches. For the baseline, the area is taken as the area of a triangle formed by the origin, the red dot and (1, 1) in the false positive rate against true positive rate plane.

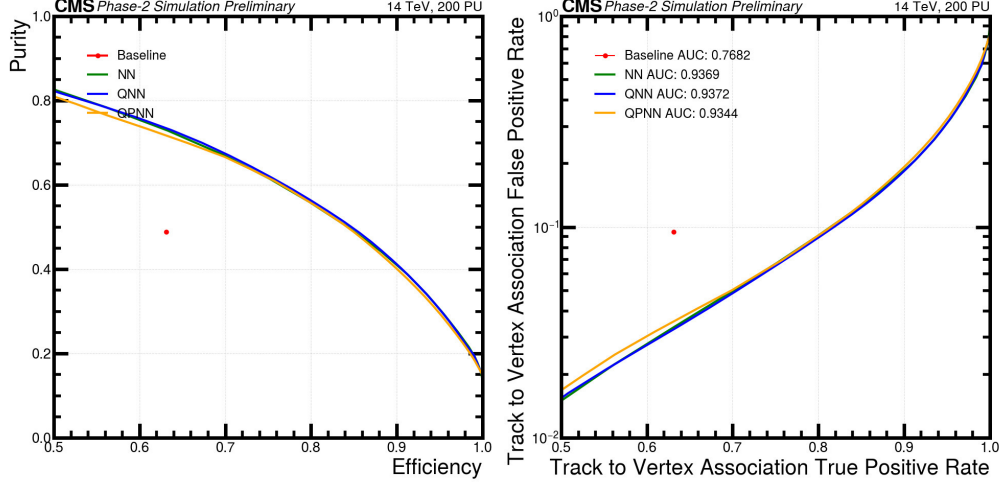


Figure 6.11: Purity against efficiency (left) and receiver operating characteristic (ROC) curves (right) for the baseline cut-based approach and the neural network track-to-vertex association with the new Kalman Filter tracks. The baseline is shown in as a red dot as it represents a single cut, as opposed to varying thresholds for the neural networks. In the ROC plot, the ROC AUC is the area under curve for the neural network approaches. For the baseline, the area is taken as the area of a triangle formed by the origin, the red dot and (1, 1) in the false positive rate against true positive rate plane.

<i>Vertex finding</i>	Latency (ns)	Initiation interval (ns)	LUTs (%)	DSP (%)	BRAMs (%)	FFs (%)
QP weight network Vitis	40	2.7	0.12	0.06	0.00	0.05
Q pattern network Vitis	35	18.9	1.53	0.00	0.00	0.92

<i>Track-to-vertex association</i>	Latency (ns)	Initiation interval (ns)	LUTs (%)	DSP (%)	BRAMs (%)	FFs (%)
QP association network Vitis	40	2.7	0.33	0.32	0.00	0.11

Table 6.1: Resource usage and latency of one SLR of a Xilinx VU13P running at 360 MHz for the quantised pruned (QP) weight network, quantised (Q) pattern network and quantised pruned (QP) association network in the end-to-end neural network vertexing. The resource usage is estimated by a Vivado implementation of the networks, while the latency and initiation interval are estimated from a simulation. Post-training quantisation is applied to reduce the resource usage.

6.3.4 Impact on downstream objects

The impact of the neural network vertexing on downstream objects is investigated. Missing transverse energy (MET) is of interest because it is a feature of many physics models beyond the Standard Model, such as Supersymmetry and dark matter models. The performance in PUPPI MET, which is MET with corrections applied to account for pileup effects [116], is first studied by only changing the vertex reconstruction algorithm.

The trigger efficiency is studied at a fixed rate of 18 kHz with the neural network vertexing, the baseline FastHisto algorithm, and truth tracks respectively. In the case of truth tracks, Monte Carlo truth-matched tracks are used as input to the FastHisto algorithm to obtain the vertex, and the truth-matched tracks are associated to the vertex. A Supersymmetric top squark sample is used for the study, as there is a large amount of MET in the sample.

As seen in Fig. 6.12 (left), the efficiency from using truth tracks comes at the top, which is expected as they use truth-matched primary vertex tracks as input and no pileup or fake tracks are associated to the vertex. The truth tracks give a perfect performance upper limit. The efficiency from neural network vertexing is found to be similar to that from the FastHisto algorithm, as shown in Fig. 6.12 (right). This is also expected as the only difference here is in the vertex position, where the biggest improvement comes from the tails of the vertex residuals. The improvement in the track-to-vertex association (which will be discussed later) should have a bigger impact on the performance compared to the vertex regression, as tracks are used in the object reconstruction.

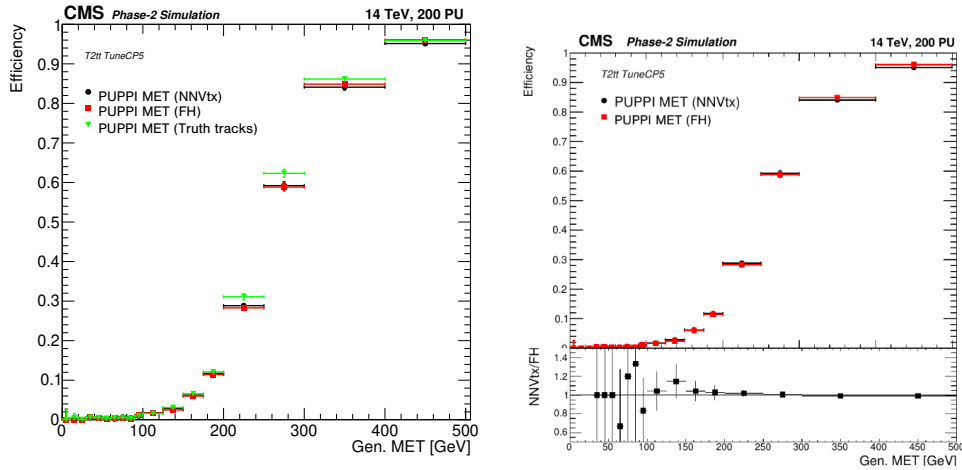


Figure 6.12: Trigger efficiency of PUPPI MET as a function of generated MET at a fixed rate of 18 kHz using the vertex from neural network vertexing, the baseline FastHisto algorithm and Monte-Carlo truth-matched tracks respectively. In the case of truth tracks, the truth-matched tracks are input to the FastHisto algorithm to obtain the vertex, and only the truth tracks are associated to the vertex. The plot on the right shows a direct comparison between the efficiencies from neural network vertexing and the FastHisto algorithm.

The trigger rate for PUPPI MET is then studied, again with only changes in vertex position. A minimum bias sample is used for the study, which consists of non-single diffractive inelastic interactions. The sample includes many different processes, and so it is suitable for the study of trigger rate. The rates are first compared between the neural network vertexing and the baseline FastHisto algorithms as a function of the offline p_T threshold, as shown in Fig. 6.13 (left). The neural network vertexing is found to give a

lower rate compared to the FastHisto algorithm for most of the p_T thresholds. For example, at 150 GeV, the rate for neural network vertexing is 343 kHz, while that for the FastHisto algorithm is 363 kHz, showing a reduction of 5.6% in rate with neural network vertexing. As the trigger efficiency is found to be similar between neural network vertexing and FastHisto with only the change in vertex position, a reduction of rate at the same efficiency shows good performance for the neural network vertexing.

The rates are also compared between the FastHisto algorithm and the truth tracks, as shown in Fig. 6.13 (right). The rate for the truth tracks is found to be lower than that for the FastHisto algorithm. This is expected as in the case of truth tracks, only Monte Carlo truth-matched tracks and no pile-up nor fake tracks are associated to the vertex.

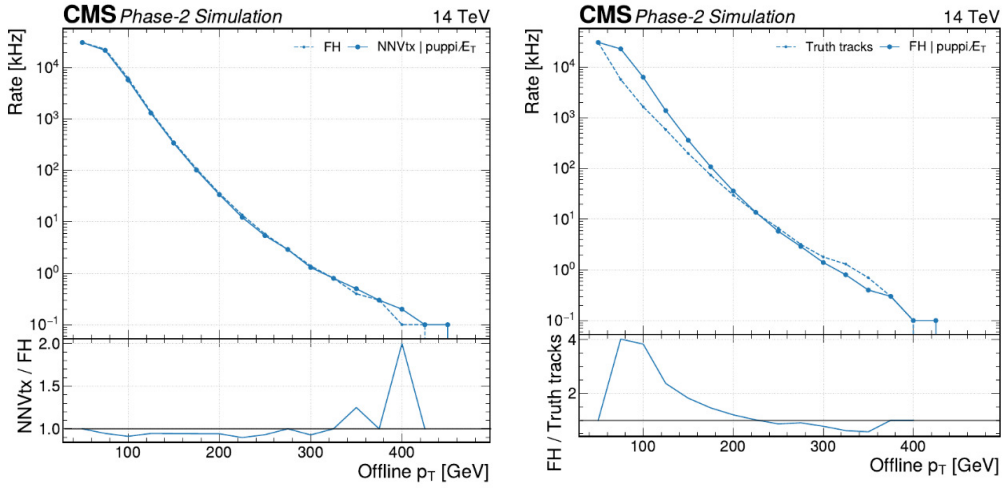


Figure 6.13: Trigger rate of PUPPI MET as a function of offline p_T threshold, making comparisons between the neural network and the FastHisto algorithm (left), and between the FastHisto and truth tracks (right). In the case of truth tracks, Monte Carlo truth-matched tracks are input to the FastHisto algorithm to obtain the vertex, and only the truth tracks are associated to the vertex. The fluctuations at lower rates are due to statistical errors.

Aside from PUPPI MET, the impact of the neural network vertexing is also studied with hadronic taus (tau leptons that decay into hadrons), which can also benefit from the improved vertexing. The neural network track-to-vertex association is implemented in addition to the vertexing in order to study the full impact of the neural network approach. The track-to-vertex association is performed within the PUPPI reconstruction algorithm as a selection on the neural network output score of the tracks. The efficiency is studied with the vector boson fusion $H \rightarrow \tau\tau$ sample, and the rate is studied with the minimum bias sample. The object under study is the neural network PUPPI tau, which uses a neural-network based reconstruction algorithm for hadronically decaying tau leptons using information from the calorimeters and tracks [108].

The matching efficiency of the neural network PUPPI tau is first studied. It is defined as

the efficiency of matching generated taus with Level-1 taus with $\Delta R < 0.1$. The performance is compared between the FastHisto algorithm and the neural network vertexing, as shown in Fig. 6.14 (left) for the barrel and Fig. 6.14 (right) for the endcap. There is a clear improvement of 10-20% in the matching efficiency when using the neural network vertexing compared to the FastHisto algorithm, especially in the barrel region.

The trigger rate for neural network PUPPI tau is then investigated. To compare the rates at fixed efficiency, different p_T cuts are applied on the Level-1 taus to obtain the same efficiency for the FastHisto and the neural network vertexing algorithms. As a reference, a p_T threshold of 30 GeV is used for the baseline FastHisto algorithm. The threshold for the neural network vertexing is tuned such that it gives the same efficiency as the FastHisto algorithm. It is found that thresholds of 33 GeV and 32 GeV for the neural network vertexing give roughly the same trigger efficiency as the FastHisto algorithm in the barrel and the endcap respectively, as shown in Fig. 6.15. The rate for single tau against the online thresholds is shown in Fig. 6.16 for the barrel and endcap respectively. It is noted that in the proposed Phase-2 trigger menu a threshold of 30 GeV is used for di-tau seeds, while the rate is shown for single tau in Fig. 6.16. In the barrel region, the rate for the FastHisto algorithm is 564 kHz at a threshold of 30 GeV, while the rate for the neural network vertexing is 514 kHz at a threshold of 33 GeV. The thresholds are chosen to compare the rates at roughly the same efficiency for the two algorithms. This shows that the rate is 8.9% lower for the neural network vertexing in the barrel, while maintaining a similar trigger efficiency as the FastHisto algorithm. In the endcap region, the rate for the FastHisto algorithm is 355 kHz at a threshold of 30 GeV, while the rate for the neural network vertexing is 350 kHz at a threshold of 32 GeV. This shows that the

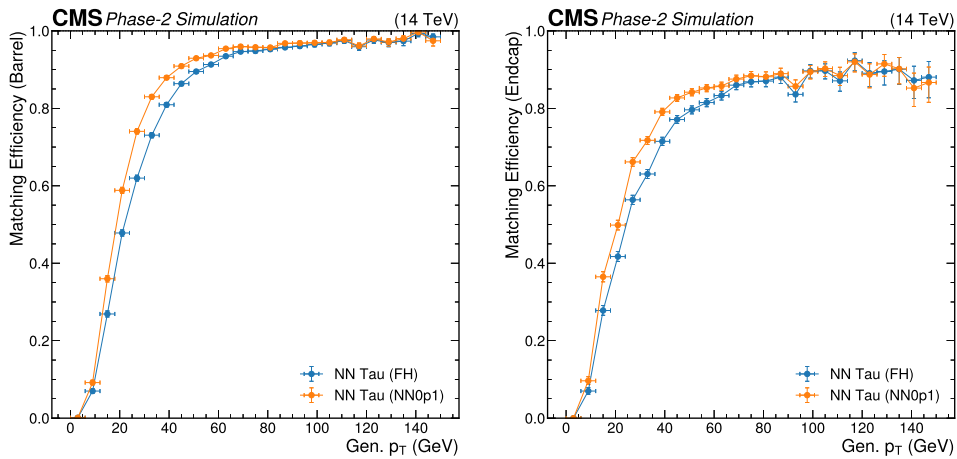


Figure 6.14: Matching efficiency of generated tau with the Level-1 neural network PUPPI tau in the barrel (left) and the endcap (right) using the FastHisto algorithm (blue) and the neural network vertexing (orange) respectively. A threshold of 0.1 is used for the association network in the neural network vertexing. A clear improvement in the matching efficiency is observed for the neural network vertexing.

rate is 1.5% lower for the neural network vertexing at a similar efficiency as the FastHisto algorithm in the endcap region.

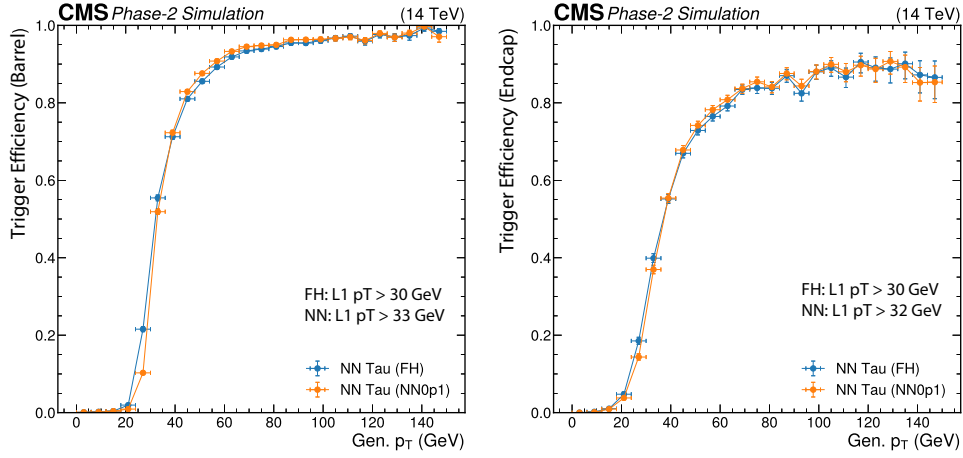


Figure 6.15: Trigger efficiency of the neural network PUPPI tau in the barrel (left) and the endcap (right) respectively. A threshold of 0.1 is used for the association network in the neural network vertexing. A Level-1 p_T threshold of 30 GeV is applied on the taus for the FastHisto algorithm, while the p_T threshold is tuned for the neural network vertexing to obtain roughly the same efficiency as the FastHisto.

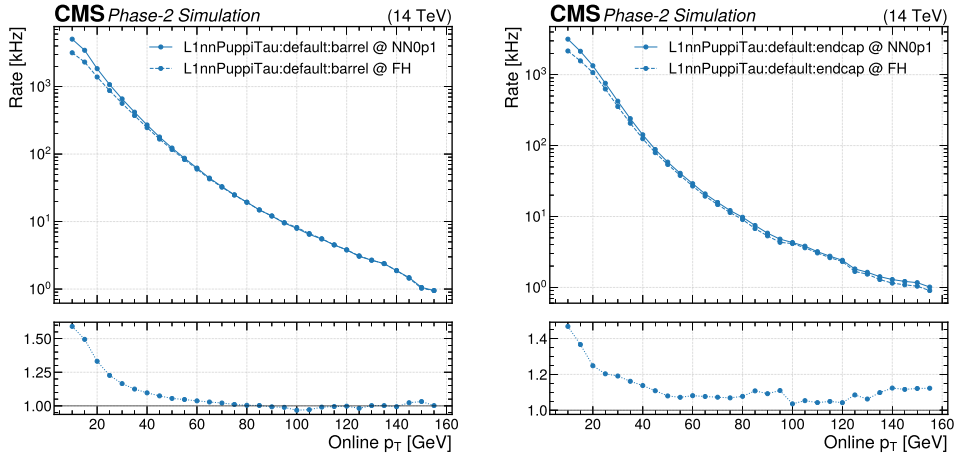


Figure 6.16: Trigger rate of the neural network PUPPI tau as a function of online threshold in the barrel (left) and the endcap (right) using the FastHisto algorithm and the neural network vertexing respectively. A threshold of 0.1 is used for the association network in the neural network vertexing.

6.3.5 Summary

The HL-LHC will provide an integrated luminosity of up to 3000 fb^{-1} , with a pileup of around 200. Although it brings great opportunities for physics searches and measurements, it also presents a challenging detector environment. The Level-1 trigger will be upgraded to include information from different subdetectors and implement sophisticated

reconstruction algorithms. In particular, the global track trigger will use tracks from the Outer Tracker to perform various tasks, including the reconstruction of the primary vertex. The baseline FastHisto algorithm uses a histogram of track z_0 weighted by p_T to find the primary vertex. Neural network vertexing is studied to improve the performance by using more complex track weightings that include for example track quality information.

The performance of the neural network vertexing is studied with the latest Monte Carlo samples of $t\bar{t}$ events. There are improvements in both the vertex regression and the track-to-vertex association for the neural network vertexing compared to the baseline FastHisto algorithm. The resource usage of the neural network implemented on a Xilinx VU13P running at 360 MHz is estimated using a Vivado implementation and a C-Simulation. Reasonable resource usage is observed for the neural network implementation.

The impact of the neural network vertexing on downstream objects is investigated for PUPPI MET and the neural network PUPPI tau. The impact of the changes in vertex position is first studied, which shows similar trigger efficiency and a reduction in rate of about 6% in the trigger rate for PUPPI MET. The neural network track-to-vertex association is then implemented and the impact is studied for the neural network PUPPI tau. An improvement of 10-20% is achieved in the matching efficiency of generated tau with the Level-1 tau. At fixed efficiency, reduction in the trigger rate of about 9% and about 2% are observed in the barrel and the endcap regions respectively.

As the neural network vertexing shows promising results, there will be continued efforts in the development of machine learning vertexing. Implementation of the neural network into firmware will also be tested to ensure matching between simulation and emulation, as well as the satisfaction of timing and resource usage requirements. Furthermore, the neural network association score could be deeply incorporated in τ algorithms in the future for more optimal use.

Chapter 7

Analysis strategy

7.1 Introduction

The search for dark showers is performed using a dataset collected in 2018 by the CMS experiment using a “data parking” strategy [117, 118], where the offline reconstruction is delayed. In the B-parking dataset the thresholds used by the trigger algorithms are lowered, facilitating the selection of low p_T muons that are displaced from the primary vertex. This dataset provides great opportunities for searching for low-mass, long-lived dark sector decays into displaced muons.

An event-level Boosted Decision Tree (BDT) is trained in the analysis to discriminate between the signal and the background. A variety of variables from different objects are used as input variables for the training. Muon variables are found to be the most discriminating variables.

In this Chapter, the B-parking dataset is first introduced, as well as the Monte Carlo samples that are produced for the analysis. The training of the BDT is then be discussed, including performance and feature studies. Finally, the event selection and event categorisation are discussed.

7.2 Datasets, triggers and Monte Carlo simulation

7.2.1 The B-parking dataset

The search targets low-mass Hidden Valley dark showers, which are hard to trigger using standard trigger thresholds. In order to maintain a manageable trigger rate, traditional triggers usually have relatively high thresholds, which in turn makes it challenging to perform searches for new particles with low mass. These final states are difficult to select

out of the huge QCD background.

One of the main challenges faced by data processing is the limited computing resources available for performing prompt reconstruction of recorded data. In standard datasets, the full offline reconstruction proceeds soon after the data are recorded, with only a short delay of about 48 hours for different detector calibration and alignment data to become available [119]. The novel data-parking strategy was designed to overcome the limitation in the ability to perform prompt reconstruction [117]. The trigger thresholds are lowered in the parking strategy, which increases the acceptance to low mass physics final states. The trigger rate increases substantially as a result, which can go beyond the capacity of the computing resources available to promptly reconstruct events as soon as they are recorded. In such case, the data stream is first transferred, unprocessed, to tape storage (the data stream is "parked"). It is kept in a raw format until sufficient computing resources become available for the events to be reconstructed, for example during shutdowns between data-taking periods. The parking strategy is illustrated in Fig. 7.1 for the 2018 data taking.

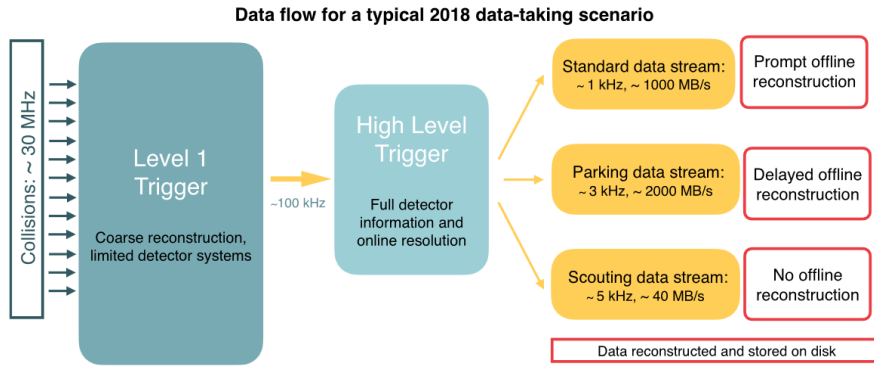


Figure 7.1: The typical Run 2 data flow during 2018, which shows the different data acquisition strategies, including the parking data stream. A value of $\mathcal{L}_{\text{inst}} = 1.2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ over a typical 2018 fill is used, which corresponds to an average pileup of 38 [117].

In the 2018 B-parking dataset, there were triggers aimed at selecting low- p_T muons that are displaced from the primary vertex. They were designed to identify muons originating from displaced b hadron decays, with kinematic requirements on the transverse momentum and the impact parameter significance of the muons. The latter is defined as $\text{IP}_{\text{sig}} = \text{IP} / \sigma_{\text{IP}}$, where IP is the transverse impact parameter of the muon and σ_{IP} is its uncertainty. It is used for reducing prompt muon background from processes such as D mesons and charmonium decays. The trigger thresholds are progressively relaxed in the Level-1 (L1) and the high-level trigger (HLT) algorithms, as the instantaneous luminosity decreases gradually in each fill. Normally the trigger rates decrease as the instantaneous luminosity decreases, which leads to the availability of spare capacity. To utilise the resources available, in the parking strategy the trigger thresholds are progressively loosened so that the trigger system operates at a rate close to its design limit. The higher trigger

rates are mitigated by delayed offline reconstruction. This led to a sample of 10^{10} b hadron decays being recorded in 2018 [117, 118].

The B-parking dataset provides great opportunities for searching for low-mass, long-lived dark sector decays into displaced muons. In the dark shower signal models, dark hadrons are produced with high multiplicity, and they undergo displaced decays into muons (either directly or in steps) with a high branching ratio. One of the muons would fire the B-parking displaced muon triggers. There are no mass restrictions and no isolation requirements in the dataset, which increase the signal acceptance. Compared to another data taking strategy known as scouting (also illustrated in Fig. 7.1), which reduces the event content for higher rate, full event information is available in the parking dataset. This is important for implementing machine learning techniques in the analysis.

At the L1 trigger, the muon is required to have either $p_T > 22$ GeV, or a looser p_T requirement ranging from 6 to 18 GeV as well as a requirement of $|\eta| < 1.5$. At the HLT, the trigger thresholds on p_T range from 7 to 12 GeV, and the thresholds on IP_{sig} vary between 3 and 6. Details of the trigger paths can be found in Appendix A.

The B-parking dataset contains 1.2×10^{10} events, and includes 1×10^{10} unbiased b hadron decays due to the imperfect trigger purity. The total luminosity recorded is 41.9 fb^{-1} . Table 7.1 shows the integrated luminosity, the mean number of pileup interactions, and the number of events recorded by each combination of L1 and HLT thresholds. The values are obtained over the periods for which each combination provided the lowest enabled L1 p_T threshold.

The pileup distribution for the dataset is given in Fig. 7.2, which shows the contributions from each of the trigger combinations. The trigger thresholds are loosened as the luminosity falls, which corresponds to lower pileup. Therefore the mean pileup decreases with lower trigger thresholds, as is evident from Table 7.1 and Fig. 7.2. The highest rates at the HLT are observed later in a fill, so a larger fraction of the data are recorded at lower pileup values compared to standard data streams.

The invariant mass distribution of oppositely charged dimuons originating from a common vertex is obtained from the B-parking dataset, as shown in Fig. 7.3. Both muons from the vertex are required to satisfy minimal kinematic and identification criteria. In addition, one of the muons from the vertex must be matched to the candidate that fires the trigger in order to obtain the vertex selected by the trigger. The contributions from each trigger combination are shown in Fig. 7.3. Peaks in the distributions which originate from the $\rho/\omega, \phi, J/\psi, \psi(2S), \Upsilon(1S)$ and Z resonances from the Standard Model are clearly visible.

L1 p_T (GeV)	HLT p_T (GeV)	HLT IP_{sig}	$\mathcal{L}_{int}(\text{fb}^{-1})$	Mean pileup	Events ($\times 10^9$)
12	12	6	8.1	37.7	0.72
10	9	6	8.4	32.9	1.67
10	9	5	1.6	33.9	0.37
9	9	6	1.6	28.2	0.34
9	9	5	5.2	28.3	1.30
9	8	5	1.6	29.2	0.52
8	9	6	1.8	24.2	0.40
8	9	5	3.8	23.9	1.00
8	8	5	1.7	24.2	0.60
8	7	4	1.5	24.5	0.84
7	8	3	0.8	19.1	0.45
7	7	4	5.5	18.6	3.56
Other combinations			0.3	-	0.12
Total			41.9	22.7	11.9

Table 7.1: Trigger configurations in the 2018 B-parking dataset, defined by unique combinations of Level-1 (L1) p_T , high-level trigger (HLT) p_T and transverse impact parameter significance (IP_{sig}) thresholds. The integrated luminosity (\mathcal{L}_{int}), the mean number of pileup interactions, and the number of events recorded by each combination are shown as the total over the periods for which each combination provided the lowest enabled L1 p_T threshold.

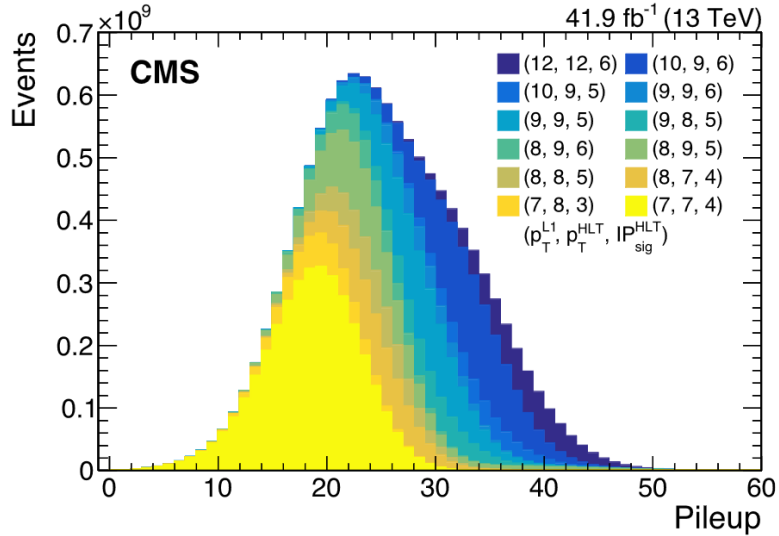


Figure 7.2: The pileup distribution for the 2018 B-parking dataset. Contributions from each individual trigger combination are shown. The areas of the histograms are normalised to the number of events recorded by each trigger.

7.2.2 Monte Carlo simulation

The signal samples are produced with PYTHIA 8.240 [53] for the vector portal signal model, and with PYTHIA 8.309 [120] for scenario A and scenario B1 models, where the signal models are defined in Section 3.2. The LHAPDF6 sets of parton distribution

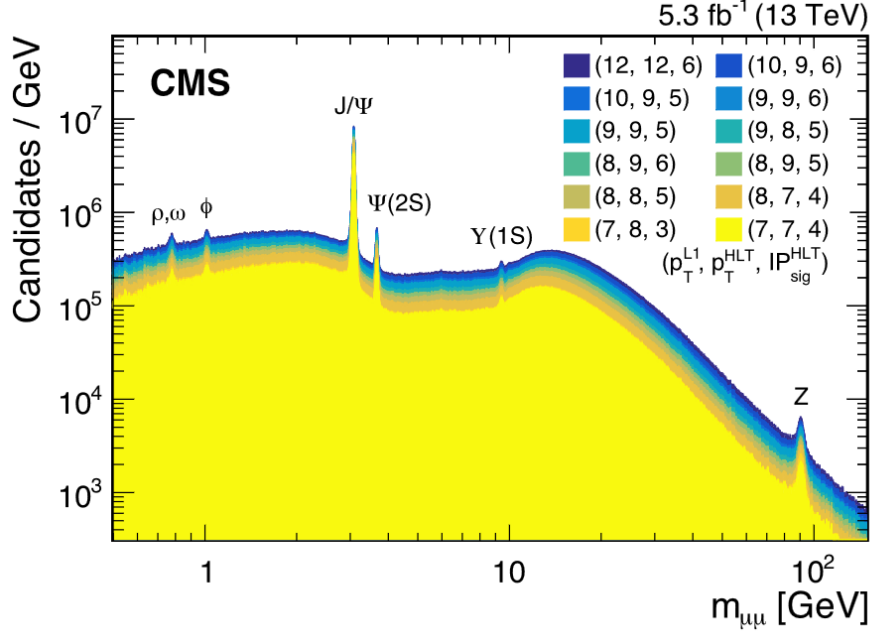


Figure 7.3: The invariant mass distribution of oppositely charged dimuons originating from a common vertex, obtained using a subset of the 2018 B-parking dataset. The contributions from each trigger combination are shown. Different Standard Model resonances are labelled.

functions [121] are used. The parameters of the models are introduced in Table 7.2. For the vector portal signal model, the parameters are the mass of the dark vector meson $m_{\tilde{\omega}}$, and its mean proper lifetime $c\tau$, which is the average lifetime in the particle’s rest frame. For scenario A and scenario B1 models, the parameters are the mass of the dark pion π_3 , the mass of the dark photon A' (which forms a resonance), and the proper lifetime of either A' (for scenario A) or π_3 (for scenario B1). The search explores a resonance mass range between 300 MeV and 20 GeV. A variety of lifetime models are considered, ranging from 0.1 mm to 500 mm. The different model points produced are summarised in Table 7.3.

The gluon-gluon fusion production mode of the Higgs is considered, which has a cross section of 43.9 pb at centre-of-mass energy of 13 TeV, calculated with a precision at NNLO+NNLL [122]. A branching ratio of 0.01 is assumed for the Higgs decaying into dark showers when the plotting the signal distributions.

Signal model	Model parameters
Vector portal	$m_{\tilde{\omega}}$: mass of the dark vector meson $c\tau$: Mean proper lifetime of the dark vector meson
Scenario A/B1	m_{π_3} : mass of the dark pion π_3 $m_{A'}$: mass of the dark photon $c\tau$: Mean proper lifetime of the dark photon (scenario A) or the dark pion π_3 (scenario B1)

Table 7.2: Model parameters of the different classes of signal models interpreted by the analysis.

	$m_{\tilde{\omega}}$ (GeV)	$c\tau$ (mm)
Vector portal	2, 5, 10, 15, 20	0.1, 0.3, 0.65, 1, 2, 4, 6.5, 10 20, 35, 50, 60, 75, 100, 200, 350, 500

	$(m_{\pi_3}$ (GeV), $m_{A'}$ (GeV))	$c\tau$ (mm)
Scenario A	(1, 0.33), (4, 0.4), (2, 0.67), (4, 1.33), (10, 1.0), (10, 3.33)	0.1, 0.25, 0.3, 0.6, 1, 2.5, 3, 6, 10, 25, 30, 60, 100 0.1, 1, 10, 100
Scenario B1	(1, 0.33), (2, 0.4), (2, 0.67), (4, 1.33)	0.1, 0.25, 0.3, 0.6, 1, 2.5, 3, 6, 10, 25, 30, 60, 100

Table 7.3: Signal models with different masses and lifetimes that are produced in the analysis.

For background Monte Carlo, a set of muon-enriched QCD background samples with a muon p_T threshold of 5 GeV is used. The samples were generated with PYTHIA 8.240. The p_T requirement is chosen so that it is below the threshold of the B parking triggers, in order to make use of the full phase space reachable by the triggers. There are twelve background samples corresponding to different p_T bins that span the full range of transverse momentum starting from 15 GeV. The cross-sections for the QCD samples are shown in Table 7.4.

QCD Monte Carlo sample	Cross-section (pb)
p_T 15-20 GeV	2799000.0
p_T 20-30 GeV	2526000.0
p_T 30-50 GeV	1362000.0
p_T 50-80 GeV	376600.0
p_T 80-120 GeV	88930.0
p_T 120-170 GeV	21230.0
p_T 170-300 GeV	7055.0
p_T 300-470 GeV	619.3
p_T 470-600 GeV	59.24
p_T 600-800 GeV	18.21
p_T 800-1000 GeV	3.275
p_T 1000-inf GeV	1.078

Table 7.4: Cross-sections for the p_T -binned QCD samples that are used in the analysis.

The trigger efficiency \times acceptance of the B-parking triggers on the different signal models are studied. Table 7.5 shows the trigger efficiency for some representative benchmark

models in the vector portal, scenario A and scenario B1 respectively. The efficiency varies from about 1% to about 30%, depending on the signal model as well as the mass and the lifetime parameters.

<i>Vector portal</i>	$m_{\tilde{\omega}} = 2 \text{ GeV}$ $c\tau = 10 \text{ mm}$	$m_{\tilde{\omega}} = 2 \text{ GeV}$ $c\tau = 100 \text{ mm}$	$m_{\tilde{\omega}} = 20 \text{ GeV}$ $c\tau = 10 \text{ mm}$	$m_{\tilde{\omega}} = 20 \text{ GeV}$ $c\tau = 100 \text{ mm}$
Trigger efficiency \times acceptance (%)	26.51	7.05	27.87	14.79

<i>Scenario A</i>	$m_{\pi 3} = 1 \text{ GeV}$, $m_{A'} = 0.33 \text{ GeV}$ $c\tau = 10 \text{ mm}$	$m_{\pi 3} = 1 \text{ GeV}$ $m_{A'} = 0.33 \text{ GeV}$ $c\tau = 100 \text{ mm}$	$m_{\pi 3} = 4 \text{ GeV}$ $m_{A'} = 1.33 \text{ GeV}$ $c\tau = 10 \text{ mm}$	$m_{\pi 3} = 4 \text{ GeV}$ $m_{A'} = 1.33 \text{ GeV}$ $c\tau = 100 \text{ mm}$
Trigger efficiency \times acceptance (%)	3.77	0.70	12.84	3.01

<i>Scenario B1</i>	$m_{\pi 3} = 1 \text{ GeV}$, $m_{A'} = 0.33 \text{ GeV}$ $c\tau = 10 \text{ mm}$	$m_{\pi 3} = 1 \text{ GeV}$ $m_{A'} = 0.33 \text{ GeV}$ $c\tau = 100 \text{ mm}$	$m_{\pi 3} = 4 \text{ GeV}$ $m_{A'} = 1.33 \text{ GeV}$ $c\tau = 10 \text{ mm}$	$m_{\pi 3} = 4 \text{ GeV}$ $m_{A'} = 1.33 \text{ GeV}$ $c\tau = 100 \text{ mm}$
Trigger efficiency \times acceptance (%)	6.35	1.27	15.66	4.84

Table 7.5: The trigger efficiency \times acceptance for representative benchmark models in the vector portal, scenario A and scenario B1.

7.3 Event-level Boosted Decision Tree

7.3.1 Introduction

In the B-parking dataset there is a large heavy flavour QCD background, so machine learning techniques are employed to provide background rejection and increase the sensitivity of the search. An event-level Boosted Decision Tree (BDT) is used to separate signal and background.

The search is performed for the displaced decay of dark showers into Standard Model particles, with a high branching ratio into muons. Therefore in the event selection at least one secondary vertex formed by two oppositely-charged muons is required. As the signal models are rich in muons, muon secondary vertices are obtained by performing vertexing using muons, which will be discussed in more detail in Section 7.4.

Minimal selections are first applied to the events, including the B parking displaced muon trigger. Also, the selection of at least one muon secondary vertex that contains a muon with $p_T > 5.0$ GeV is applied. Features that discriminate between signal and background are then studied. Some example features are shown in Fig. 7.4.

It can be seen from Fig. 7.4 that there is a higher number of muon secondary vertices and a much higher number of muons in the signal events compared to the background events. The high muon multiplicity in signal comes from the high multiplicity of dark mesons in the events, up to an average of about five dark mesons per event in the vector portal for example. There is also a high branching ratio for the dark mesons to eventually decay into muons. Moreover, muon variables are found to provide excellent discrimination between signal and background. For example, the signal muons have a higher transverse impact parameter (d_{xy}) compared to the background muons, as expected since the signal models contain a large number of displaced decays that produce muons. Furthermore, the impact parameter significance of the signal muons are higher than that of the background muons, which shows that the trigger is suited for selecting the signal events as it imposes a requirement on the impact parameter significance. Variables of muons and muon secondary vertices are included as input variables to the BDT as they have high discriminating power.

In addition, variables from secondary vertices are included so that the BDT can maintain good performance in searching for displaced decays in general. The full list of variables is presented in Appendix B. The masses of secondary vertices and muon secondary vertices are not used as input variables as they might lead to sculpting of the dimuon mass distribution, which is used for the fit in the analysis.

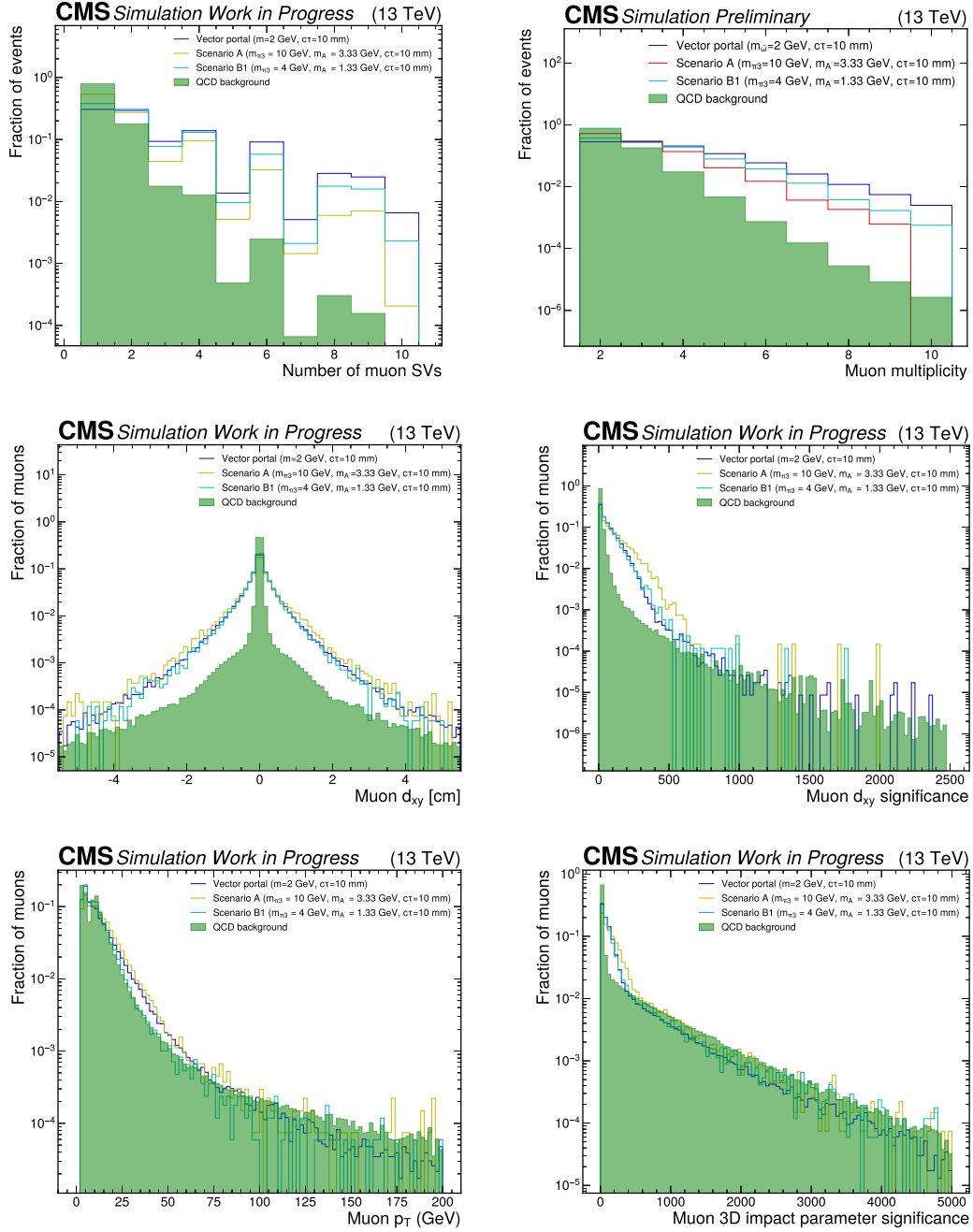


Figure 7.4: Distributions of some variables that are used in the BDT training for the QCD background and benchmark signal models.

A maximum of 8 muons, 8 secondary vertices and 8 muon secondary vertices are used in each event, and the input variables are filled with an arbitrary value (-999.0) when the number of objects in the event is below the maximum. This results in a maximum of about 350 input variables to the BDT. However by considering the average number of each object in the event, the average number of input variables is around 110. When choosing the sets of objects to use, the muons are ordered in decreasing p_T , the secondary vertices are ordered in decreasing decay length, and the muon secondary vertices are ordered in

increasing χ^2 . The secondary vertices in the signal are more displaced than those in the background, so including the most displaced secondary vertices in each event can increase the discriminating power of the BDT. In the analysis the muon secondary vertex with the minimum χ^2 is used in a mass fit to extract the signal, so the muon secondary vertices are ordered in χ^2 for the BDT.

Different machine learning models have been studied for distinguishing signal from background using Monte Carlo samples. Figure 7.5 shows the performance of different models and cut-based selections in terms of the true positive rate against the false positive rate, which are represented by the receiver operating characteristic (ROC) curves. They can also be interpreted as showing the signal efficiency against background efficiency for each model. The XGBoost [123] BDT model is found to have better performance than other benchmark neural network models and cut-based selections, as seen by the larger area under its ROC curve. Therefore, the XGBoost model is used in the analysis. With the sizes of the Monte Carlo samples available, XGBoost is found to be the best of the implemented models. It is robust against the input samples, with little change in performance for the different signal models. It is also fast and easy to deploy, which allows the machine learning model to be trained efficiently without the need of a huge amount of computing resources.

The BDT is trained with mostly the XGBoost default hyperparameters, with custom changes to a few. For example, the number of boost rounds is tuned according to the size of the training samples. The hyperparameters that are used for the BDT training can be found in Appendix C. A stable training setup is achieved. When performing the training, the signal and background events are carefully weighted to give equal contribution to the training by using an event-by-event weighted binary cross entropy loss. One BDT is trained per signal model, using all model points of the model. All signal model point samples (N samples) are reweighted to give equal contribution ($1/N$) to the total signal class, after trigger and pre-selection cuts are applied. Over-training is avoided by monitoring the validation loss and validation AUC per iteration.

Conditional training (also known as parametric training) using the mass and lifetime parameters of the model is also tested. Due to the limited sizes of the Monte Carlo training samples, it is found to give similar results to the BDT model that is unconditional on the signal model parameters. The BDT training is performed using ICENET [124].

For each class of signal model, a BDT is trained with the signal and background Monte Carlo samples. Details of the training datasets are summarised in Table 7.6.

The loss and the ROC AUC against the epochs for the training and the validation datasets are shown in Fig. 7.6, using the training with the vector portal model as an example. It

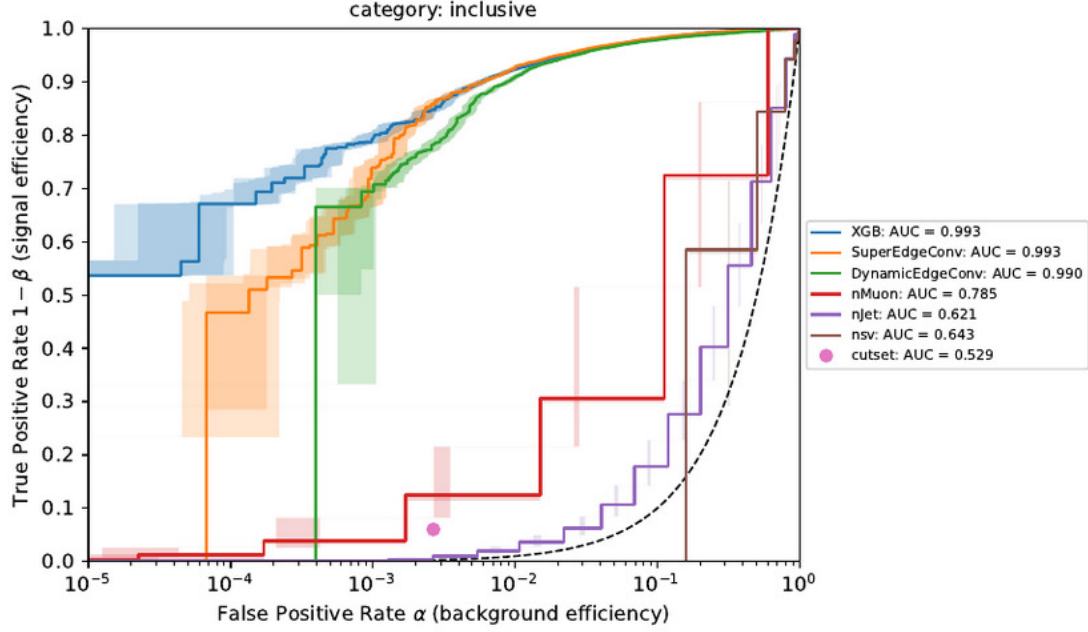


Figure 7.5: The receiver operating characteristic (ROC) curves of a range of machine learning techniques. The plot compares the performance of the different machine learning models to that of cut-based selections (which require at least one secondary vertex, at least one jet and at least four muons). The XGB (XGBoost) model is found to give the best performance, as indicated by the largest area under its ROC curve. The error bands show the statistical uncertainties in both the horizontal (false positive rate) and vertical (true positive rate) directions. The x-axis uses a logarithmic scale, and sometimes the error bands from neighbouring bins overlap with each other. The error bands are plotted at the bin centres, and they are shifted due to the log-x axis. The dashed line shows the performance for a random guess as a reference [124].

can be seen that both the training and the validation losses decrease through the epochs, with the validation loss flattening gradually. It can also be seen that the training and the validation AUC increase throughout the training. Therefore no overtraining is observed.

The number of epochs, which is one of the hyperparameters, is tuned according to the size of the training samples, such that there is good performance and no overtraining. With the available time and computing resources, some of the other hyperparameters are fine-tuned to optimise the performance, while others are kept as the XGBoost default values.

7.3.2 Study of jet variables as input variables

Studies are performed to investigate the impact of jet variables on the performance of the BDT. Jet variables are added to the set of input variables that are used for the training, with a maximum of 8 jets used per event. The list of jet variables used for the studies can be found in Appendix B.

A BDT with the additional jet variables is trained for each of the vector portal, scenario

Samples	Number of events used for the BDT
Vector portal	240,333
Scenario A	13,660
Scenario B1	33,210
QCD background	1,464,720

Table 7.6: Training samples for the event-level BDT. A BDT is trained for each class of signal model. The number of events passing pre-selections that are used for the BDT is shown for each signal model. The training/testing/validation split is 60:30:10.

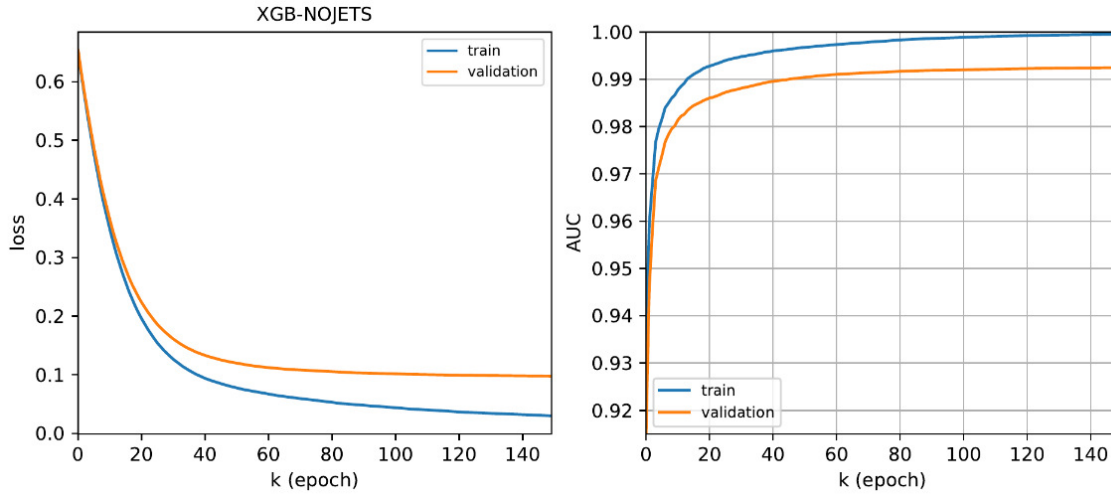


Figure 7.6: The loss (left) and the ROC AUC (right) against the epochs for the vector portal training. They are shown for both the training and the validation datasets.

A and scenario B1 model. In each case, the performance is compared to the original BDT model, which is trained with variables from muons, secondary vertices and muon secondary vertices. The ROC curves are shown for each model in Fig. 7.7.

Comparing the performance of the BDT with and without the jet variables, it can be seen from Fig. 7.7 that the difference in the AUC of the ROC curves is only about 0.5% in the different signal models. This shows that the jet variables make a minimal difference to the performance of the BDT. Therefore it is sensible to exclude the jet variables as input to simplify the BDT model.

More importantly, removing jet variables makes the BDT more model-independent, and so it is easier to study a wider range of signal models of interest, other than the Hidden Valley models. Therefore the jet variables are removed so that the BDT can be more model-agnostic and can be recast for other searches more easily.

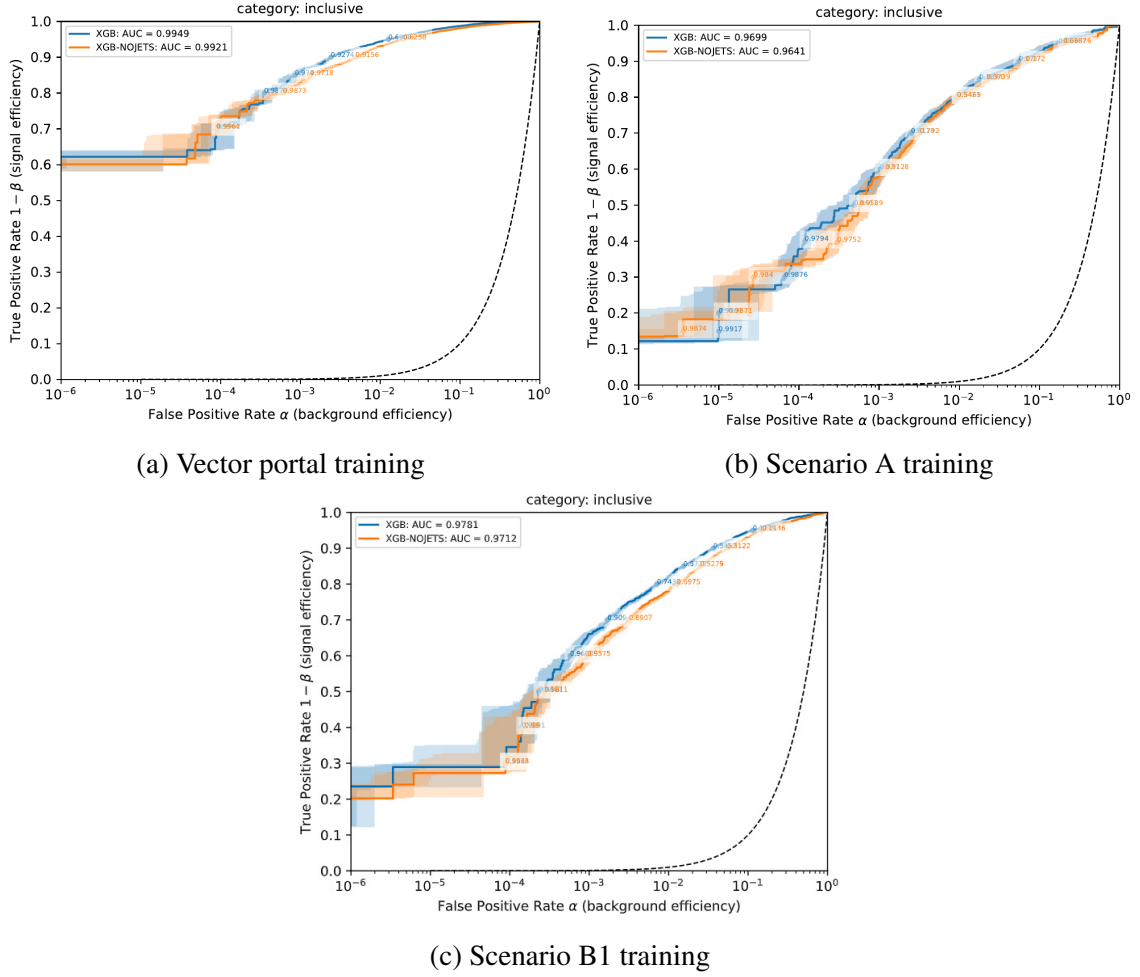


Figure 7.7: ROC curves showing the performances of the BDT trained with variables from muons, secondary vertices and muon secondary vertices (XGB-NOJETS) compared to the BDT trained with additional jet variables (XGB). The BDTs are trained with the vector portal, scenario A and scenario B1 signal model respectively.

7.3.3 Training results

The BDT output for the signal and the background are shown in Fig. 7.8 for the training performed for the vector portal, scenario A and scenario B1 model respectively. The BDT output for each event ranges from zero to one, which gives the probability of the event being a signal event based on the numerous variables used as input. As seen from Fig. 7.8, the BDT output distribution for the signal peaks at one while the distribution for the background peaks at zero. There is a clear separation between the BDT output distributions for the signal and the background, which shows the performance of the BDT in discriminating between the two.

The BDT is used for suppressing background, which is performed by applying a cut on

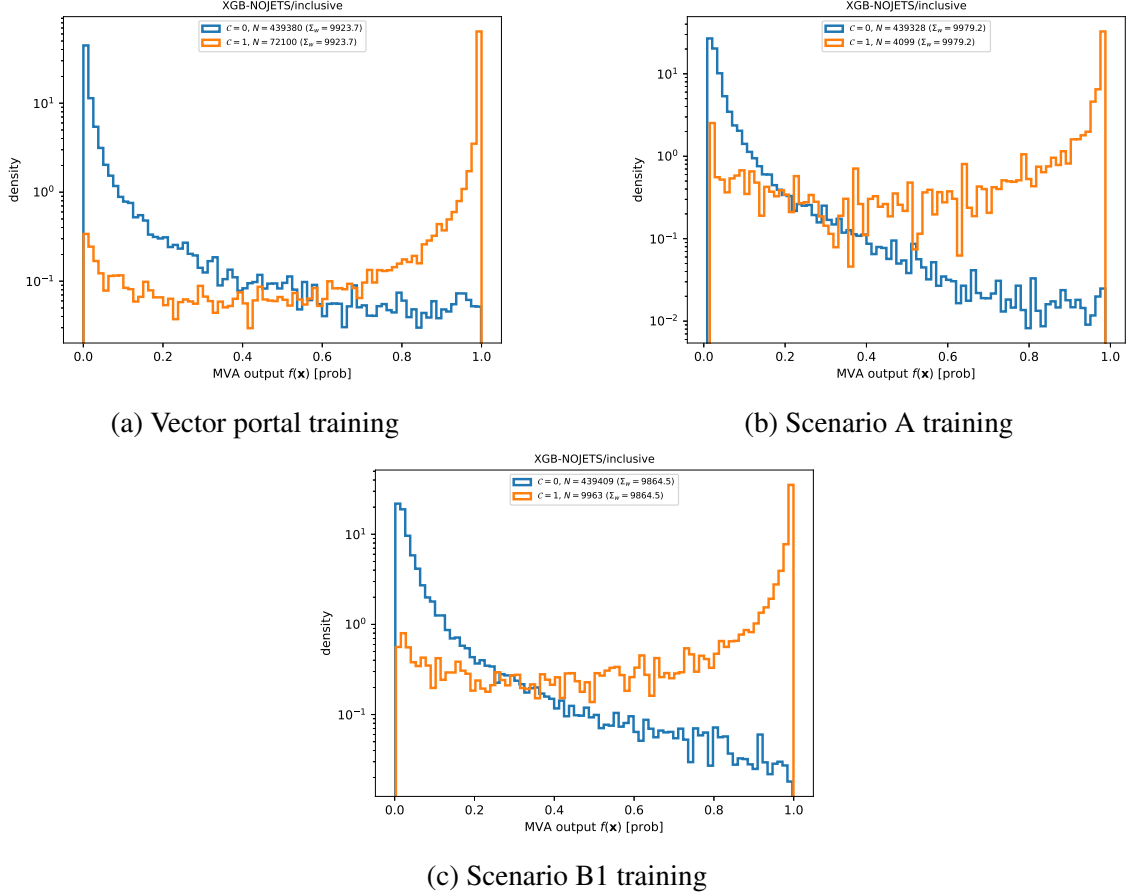


Figure 7.8: BDT output distributions for the signal and the background from the training performed with the vector portal, scenario A and scenario B1 model respectively. The BDTs use variables from muons, secondary vertices and muon secondary vertices. $C = 0$ denotes the distribution for the QCD background sample, while $C = 1$ denotes the distribution for the signal sample.

the BDT output. The ROC curves shown in Fig. 7.7 show the signal efficiency against the background efficiency at different BDT thresholds. A threshold is found to give low background efficiency and high signal efficiency in order to achieve good sensitivity. Fig. 7.7 shows that as the background efficiency goes down to 10^{-4} or further, the signal efficiency remains high, at about 30 - 60% depending on the signal model. Due to the limited sizes of the training samples, the uncertainties on the signal and background efficiencies grow as the background efficiency goes down, as indicated by the shaded region around the ROC curves in Fig. 7.7. Therefore, an initial working point is chosen for the BDT that gives 10^{-4} background efficiency with higher than 30% signal efficiency. This working point will be tuned later when the BDT is deployed on data, which contain a larger number of events.

The BDT uses a wide range of features from muons, secondary vertices and muon secondary vertices, as shown in Appendix B. The importance of each input feature is studied by producing the feature importance plot for the training. Fig. 7.9 shows the feature

importance of the vector portal training in terms of the "gain" score, which shows the relative importance of each feature for each tree in the BDT model when making a prediction. The top 50 features are shown in Fig. 7.9, which are features with the highest discriminating power. It is found that the muon-related variables make the biggest contribution to the performance of the BDT, with nearly all variables in the top 50 coming from muons and muon secondary vertices. This is expected as the signal models are rich in displaced muons originating from the decay of dark showers.

The distributions of some examples of the most discriminating features are shown in Fig. 7.10 and Fig. 7.11 for the signal, the background and a small sample of data. The χ^2 of the muon secondary vertices is found to be one of the most discriminating variables. The signal muon secondary vertices are better fitted and have a lower χ^2 compared to the background, which makes sense because the background is dominated by muon vertices from combinatorics. Moreover, the transverse displacement and transverse displacement significance of the muon secondary vertices are found to be strong variables for discriminating between the signal and the background. This is expected as the signal models feature dark hadrons of high multiplicity decaying into muons with a high branching ratio, producing muon secondary vertices with higher displacement and displacement significance compared to those in the QCD background, which are mostly produced by prompt processes.

The impact parameter of the muons is also found to have good discriminating power, with the signal muons having a higher impact parameter than the background muons as the signal muons are more displaced. The signal muons are also less likely to pass the tight identification, which has requirements on the transverse impact parameter ($d_{xy} < 2$ mm) and the longitudinal distance ($d_z < 5$ mm) of the tracker track. Moreover, the signal in the vector portal model is more pointing than the background, which will be discussed in more detail in Section 7.5. Furthermore, the signal muons are found to have higher p_T and lower relative isolation, which is defined as the sum of the energy inside a cone of radius $\Delta R = 0.3$ relative to the muon transverse momentum [125].

The BDT is trained with the signal and the QCD background Monte Carlo samples. The background sample contains about 1.5 million events and it is signal free. As the BDT will be deployed on data to perform the search, it is important to ensure that the input features used for the training are well-modelled in simulation compared to data. As seen from Fig. 7.10 and Fig. 7.11, there is good agreement between the small sample of data and the background Monte Carlo in general in the input features used for the BDT. A direct comparison between data and Monte Carlo for some example features are shown in Fig. 7.12. The data to Monte Carlo ratio is found to be close to unity for the features, which shows good agreement between data and simulation.

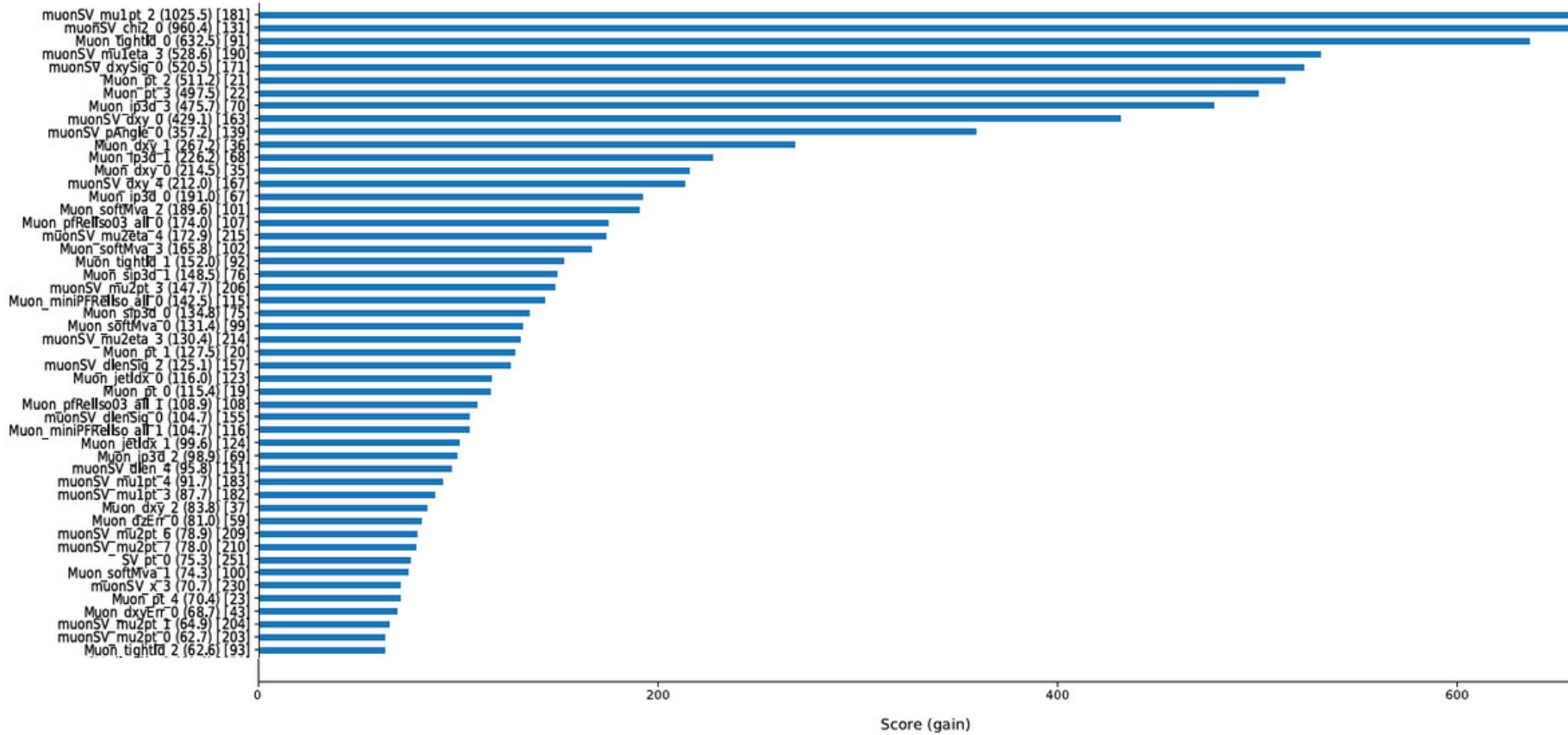


Figure 7.9: Feature importance plot for the vector portal training showing the top 50 features that have the highest "gain" scores [126], which are features that bring the most improvement to the performance of the BDT when it makes selections on those features. The numbers in round brackets show the feature importance scores for each feature. The x -axis is truncated for readability. "muonSV" refers to muon secondary vertex. The number at the end of the feature name corresponds to the index of the object.

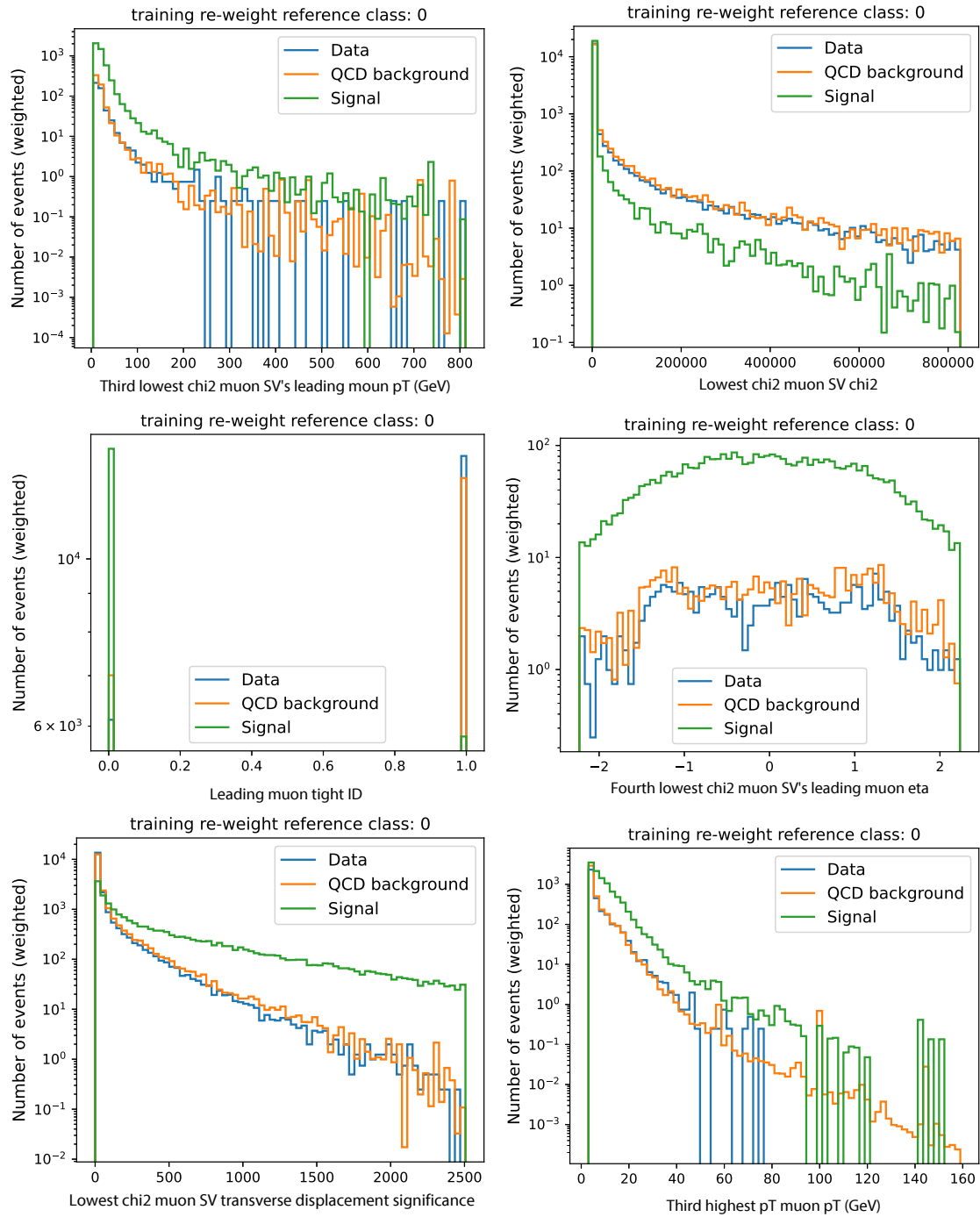


Figure 7.10: Distributions of example features which provide the best discriminating power between the signal and the background. The distributions are shown for the vector portal signal, the QCD background and a small subset of data. The signal and the data are reweighted to give the same number of weighted events as the QCD background. The plots are normalised to the same number of events, but not necessarily the same number of objects.

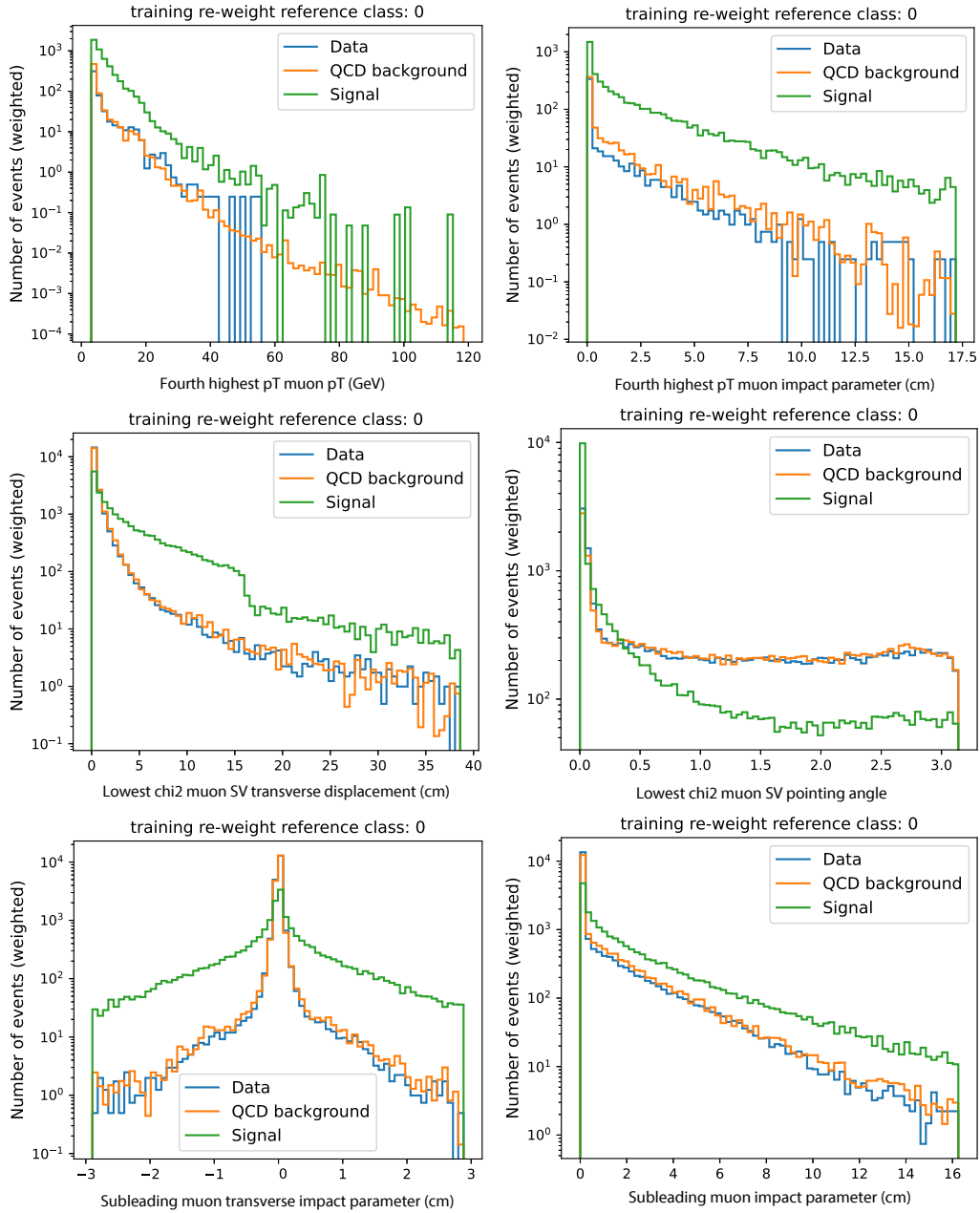


Figure 7.11: Distributions of example features which provide the best discriminating power between the signal and the background. The distributions are shown for the vector portal signal, the QCD background and a small subset of data. The signal and the data are reweighted to give the same number of weighted events as the QCD background. The plots are normalised to the same number of events, but not necessarily the same number of objects.

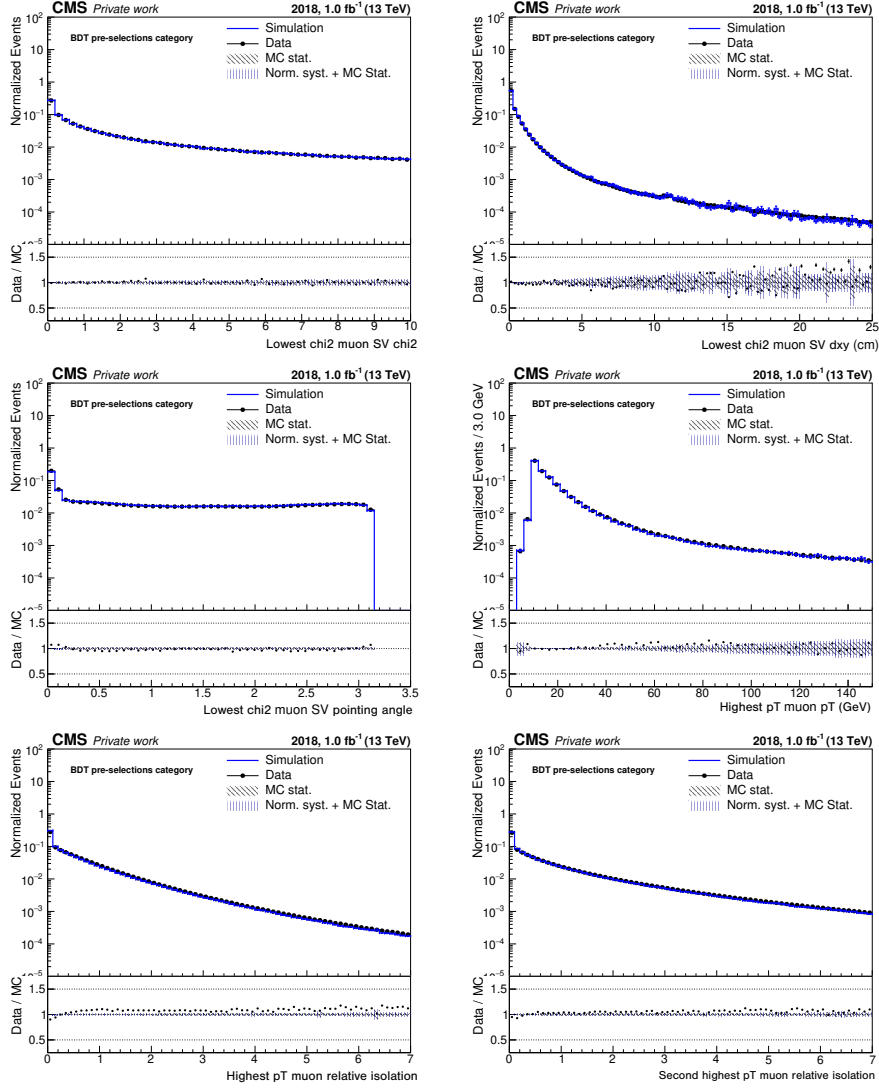


Figure 7.12: Data versus Monte Carlo distributions of example features which provide great discriminating power between the signal and the background.

7.3.4 Study of missing transverse energy input variables

The signal models involve the production of dark hadrons formed by dark QCD, some of which are stable in the detector and do not decay back to Standard Model particles. As dark hadrons do not interact with the detector, the stable dark hadrons will escape the detector as MET.

MET variables are studied to look at the impact on the performance of the BDT. The distributions of MET p_T and ϕ are shown in Fig. 7.13 for the signal, background simulation and a small subset of data.

As seen in Fig. 7.13, the MET variables provide some discrimination between the signal and the background. In particular, the signal models have a higher MET p_T compared to the background. This is expected as there is more genuine MET from dark hadrons

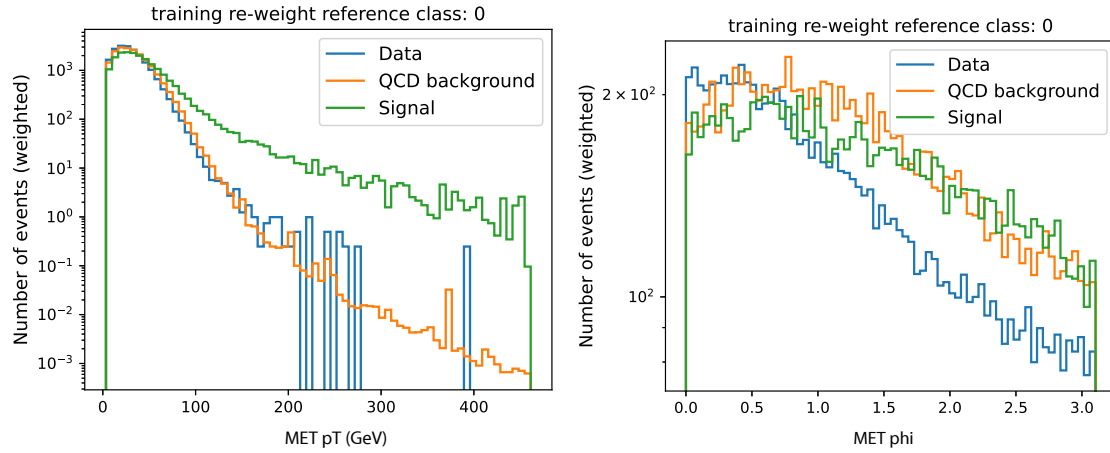


Figure 7.13: Distributions of MET variables. The distributions are shown for the vector portal signal, the QCD background and a small subset of data. The signal and the data are reweighted to give the same number of weighted events as the QCD background.

in the signal models compared to the Standard Model QCD background. However, the agreement between data and simulation is not good for the MET variables, especially for the MET ϕ distribution.

The performance of the BDT with the additional MET variables is studied for trainings with different signal models. Fig. 7.14 shows the ROC curves for the BDT with and without the additional MET variables. It is found that they only bring a small improvement to the performance of the BDT, as the increase in the ROC AUC is less than 1% in most cases after the addition of the MET variables as input variables. It is noted that there is some improvement in the signal efficiency (for example at the 10^{-4} background efficiency working point) when MET variables are added for scenario B1, which is due to the higher MET p_T in the model points considered. However, the MET variables are not well modelled, and using them as input variables will also make the BDT less general and less applicable to other models.

As the BDT is trained with simulations but deployed on data in the analysis, it is important to ensure that the variables used in simulations model those in the data well, which is not the case for the MET variables. As the MET variables are not well modelled and do not bring significant improvement to the performance of the BDT in general, they are not added to the set of input variables for the BDT used in the analysis.

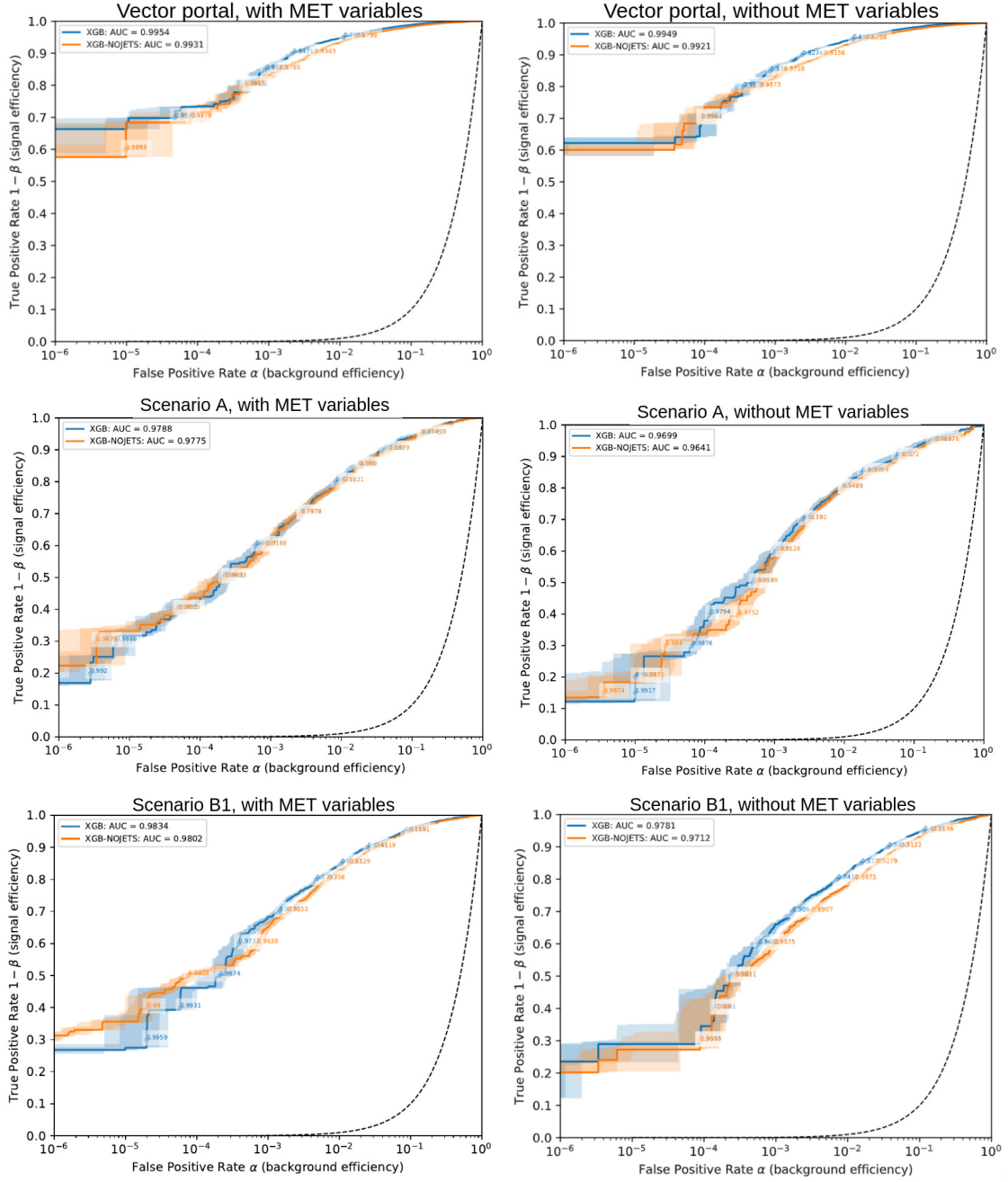


Figure 7.14: ROC curves showing the performances of the BDTs trained with or without variables of missing transverse energy. The trainings of both BDTs are performed with the vector portal, scenario A and scenario B1 signal models respectively.

7.3.5 Study of spatial relationships of muon secondary vertices as input variables

The dark shower signal models contain dark hadrons with high multiplicity that decay into pairs of muons with a high branching ratio. The spatial relationships between pairs of muon secondary vertices are studied to see if for example the vertices are more collimated in the signal compared to the background.

In the BDT training a wide range of variables from muon secondary vertices have been used as input variables, which are shown in Fig. 7.9 to give strong discriminating power between the signal and the background. The input variables include the x, y and z positions of the vertex, but they do not readily provide information about the spatial relationships between the vertices. For this study the spatial separations between vertices are measured using ΔR . The spatial separations are then ordered from lowest to highest, and the first 28 values are used as input variables to the BDT. Some examples of the ΔR distributions are shown in Fig. 7.15, which correspond to the four smallest ΔR between pairs of muon secondary vertices in each event. The distributions are shown for the vector portal signal model, the QCD background and a small subset of data. The muon secondary vertices are found to be slightly more collimated in the signal compared to the background. For example, the minimum ΔR between the vertices is about 40% smaller in the signal compared to the background. However, there are noticeable discrepancies between the background simulation and the data in the ΔR distributions, which show that the simulations of the angular separations between the muon secondary vertices are not well modelled.

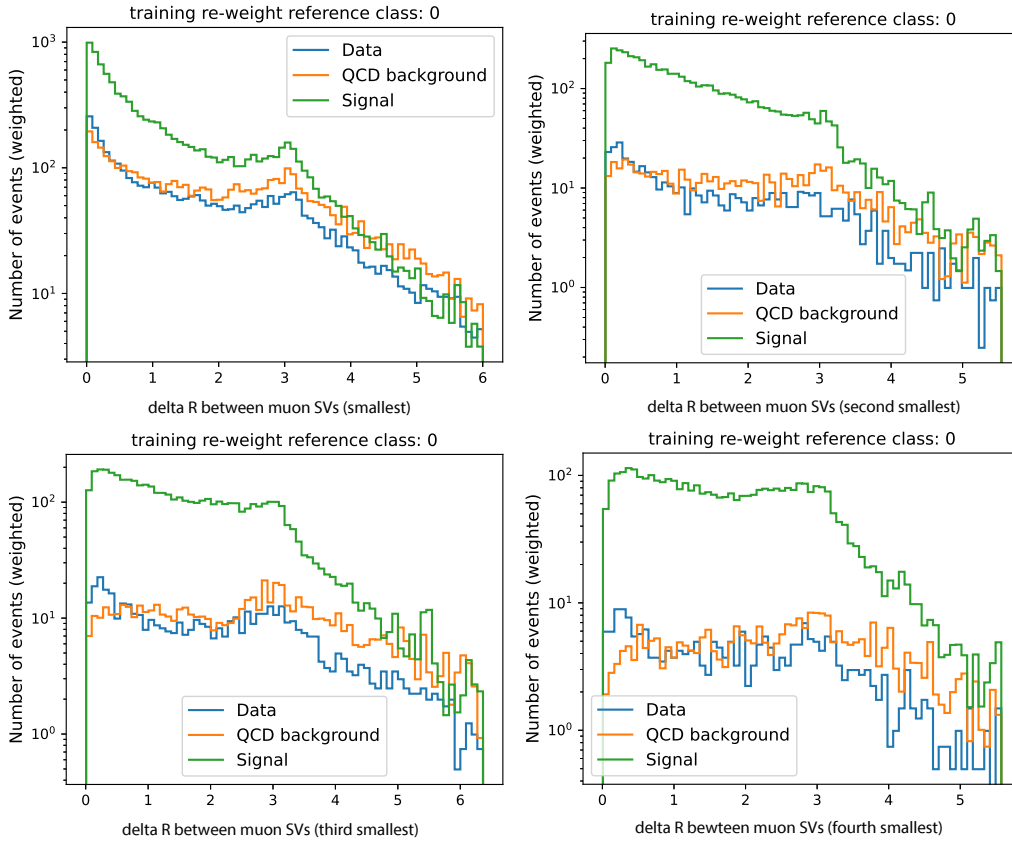


Figure 7.15: Distributions of ΔR between muon secondary vertices. The plots show the four smallest ΔR between pairs of muon secondary vertices in each event. The distributions are shown for the vector portal signal, the QCD background and a small subset of data. The signal and the data are reweighted to give the same number of weighted events as the QCD background.

The BDT is trained with the additional ΔR variables between the muon secondary vertices

for each of the vector portal, scenario A and scenario B1 models. The ROC curves of the BDTs with and without the ΔR variables are shown in Fig. 7.16. The plots show essentially the same performance in the BDT after the addition of the ΔR between the muon secondary vertices as input variables, with the change in ROC AUC to be within 1%. This shows that the spatial relationships between the muon secondary vertices bring negligible improvement to the performance of the BDT. As discussed in Section 7.3.3, other variables of the muon secondary vertices, particularly the χ^2 and the displacements, provide much greater discriminating power between the signal and the background.

As the BDT is trained with simulations but deployed on data in the analysis, it is important to ensure that the variables used in simulations model those in the data well, which is not the case for the ΔR variables. As the spatial separations between the muon secondary vertices bring negligible improvement to the performance of the BDT, and the variables are not well modelled, the ΔR between muon secondary vertices are not used as input variables for the BDT in the analysis.

7.3.6 Summary

As there is a substantial QCD background in the B-parking dataset in which the search is performed, machine learning techniques are studied to distinguish signal from background. An event-level BDT is found to give strong performance in discriminating between signal and background, with the sizes of the Monte Carlo training samples and the computing resources available.

A wide variety of event features are used as input variables to the BDT. Muon variables are found to be excellent variables with high discriminating power. This is due to the large number of displaced muons from the decay of dark showers in the signal models. Therefore variables from muons and muon secondary vertices are used as input variables. In order to maintain high performance of the BDT for searches of displaced decays in general, variables of secondary vertices are also included. Variables from jets, MET and the spatial separations between muon secondary vertices are found to provide limited improvement to the performance of the BDT, so they are not used as input variables.

An initial working point is chosen for the BDT that gives 10^{-4} background efficiency with higher than 30% signal efficiency, which shows the capability of the BDT.

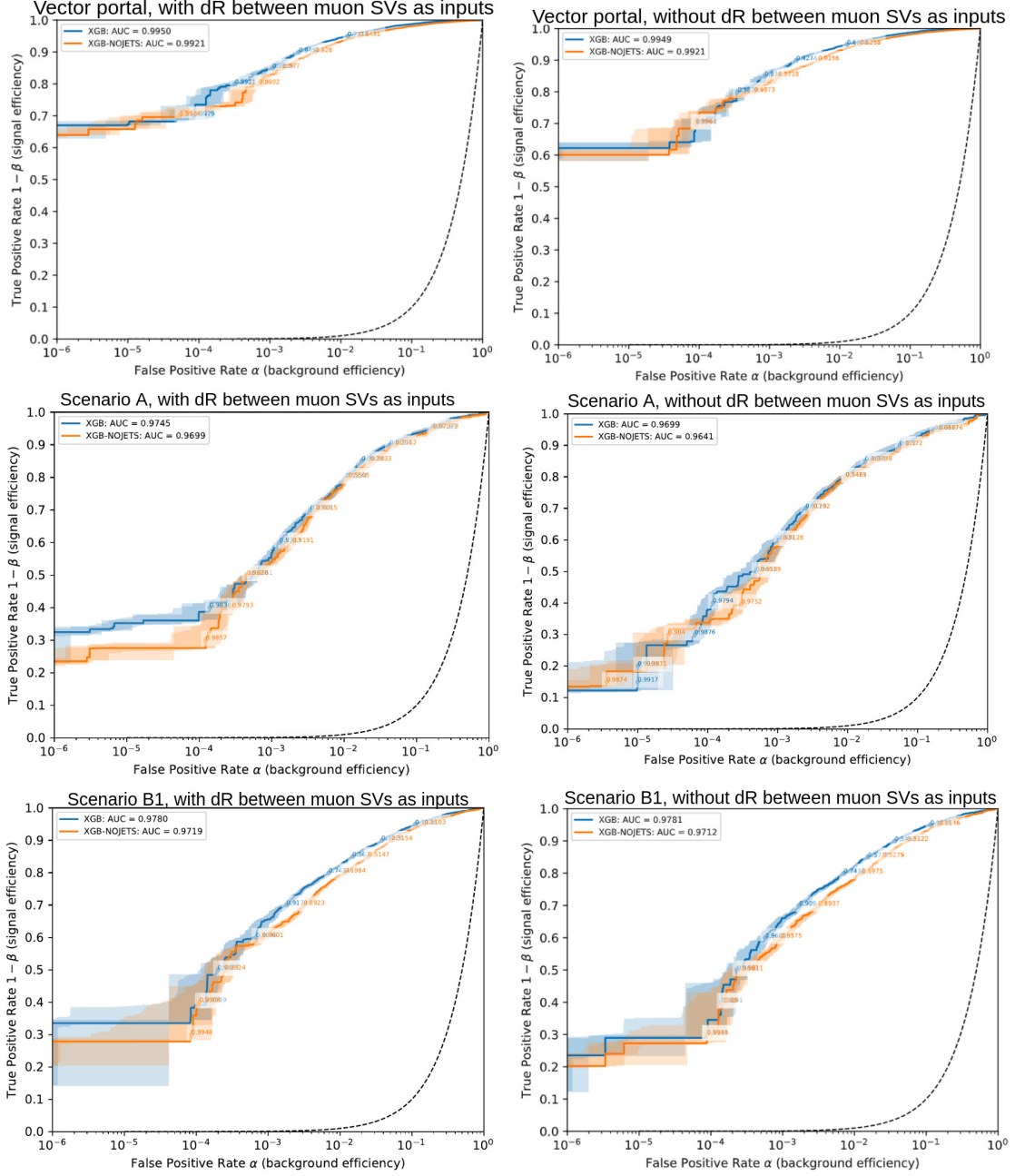


Figure 7.16: ROC curves showing the performances of the BDTs trained with or without input variables from the ΔR between pairs of muon secondary vertices. The trainings of both BDTs are performed with the vector portal, scenario A and scenario B1 signal models respectively.

7.4 Event selection

Events are first selected by using the trigger paths detailed in Section 7.2.1. All triggers of the B-parking dataset are included in order to exploit the full integrated luminosity of the dataset for the search, maximising the sensitivity. Trigger matching is then performed to ensure there is a well-defined efficiency. The reconstructed muon is matched to the

HLT muon (which is the trigger object) by requiring $\Delta R < 0.1$. Offline selections are then applied to the muon to reinforce the trigger matching. The triggering muon is required to satisfy $p_T > 9$ GeV, $IP_{\text{sig}} > 6$ and $|\eta| < 1.5$. The selections are made in the turn-on regions of some triggers of the dataset (e.g. for HLT_Mu9_IP6), which is beneficial for the acceptance. To account for potential differences in the trigger efficiencies in data and Monte Carlo, studies are performed which will be discussed in Section 8.6.4. Scale factors are applied to the Monte Carlo (also discussed in Section 8.6.4) to account for any mis-modelling of the trigger efficiency.

Moreover, the triggering muon is required to pass the loose muon identification as a quality requirement. A loose muon is defined as a Particle Flow muon that is either a global or a tracker muon [127, 128]. This requires the muon candidate's track in the muon system to be matched to a track in the inner tracking system. For global muons, a search for tracker tracks in the inner tracking system that can be matched is performed for each standalone muon track in the muon system. The best-matching tracker track is selected, and then a track fit using the hits of the selected pair of tracker track and standalone muon is performed using the Kalman Filter technique [97]. For tracker muons, each tracker track that passes loose requirements are extrapolated to the muon stations to see if there is any matched segment in the muon stations. The loose identification does not impose requirements on the muon's impact parameter [88], so it is suitable for studying displaced muons in the analysis.

After applying the above baseline selections, a selection is made on the BDT output score of the events. The distribution of BDT scores for the QCD background simulation and two illustrative signal models in the vector portal are shown in Fig. 7.17 (left). As an initial working point, a threshold of the BDT that gives 10^4 background rejection is chosen. The corresponding thresholds for the different BDTs trained for the different physics models are found to be 0.997, 0.98 and 0.991 for the vector portal, scenario A and scenario B1 models respectively. Figure 7.17 (right) shows the selection efficiency against $(1 - \text{BDT score})$ for the QCD background simulation and two illustrative signal models in the vector portal. It can be seen that for the working point of 10^{-4} background efficiency, the signal efficiency is more than 30% depending on the signal model. The initial BDT working point shows good performance in selection efficiencies, and it will be optimised in the final fit using data control regions, which will be described in Section 8.5.

At least one secondary vertex (SV) formed by a pair of oppositely charged (opposite-sign, OS) muons is also required. The vertexing is performed using the Kalman Filter algorithm [97] on muons which pass loose kinematic selections. The muons in the SV are required to have either $p_T > 15$ GeV or $p_T > 3$ GeV with muon identification requirements. The selections are loose in order to cover a larger kinematic phase space for the displaced

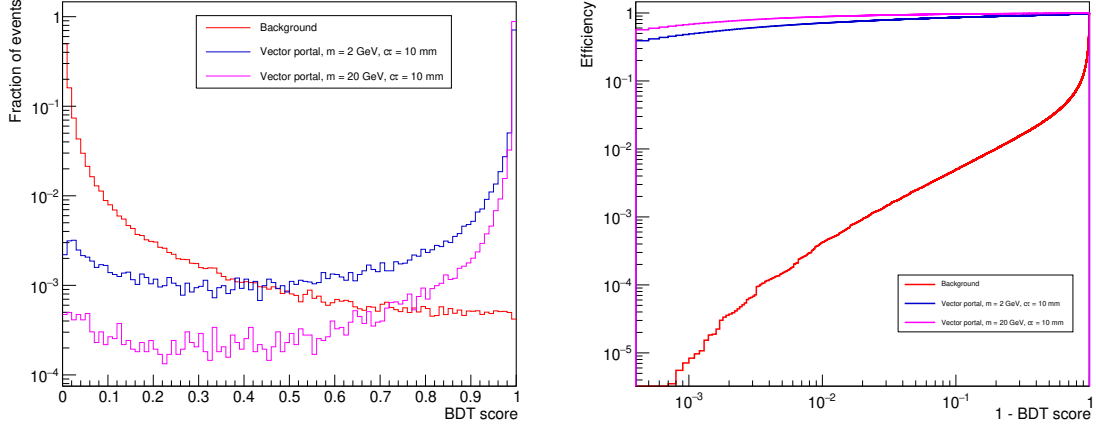


Figure 7.17: Distribution of BDT score (left) and BDT selection efficiency (right) for the QCD background and selected signal models.

muons. Also, the range of the vertexing algorithm is extended to improve the efficiency for decays outside the tracker.

For events passing selections with more than one muon SV, only the muon SV with the lowest χ^2 is used to determine the parameters for categorisation and fitting. Other choices of muon SV have been investigated, which are the muon SV with the highest p_T , the highest d_{xy} or the highest d_{xy} significance. Worse sensitivity is observed with these choices compared to using the muon SV with the lowest χ^2 . Therefore, the muon SV with the lowest χ^2 is used for the mass fit to give the best sensitivity.

The muon SV is also required to satisfy quality criteria, including $\Delta R(\mu, \mu) < 1.2$, in order to reduce fakes and combinatorics particularly in the background. $\Delta R(\mu, \mu)$ corresponds to the ΔR between the two muons in the muon SV. Distribution of $\Delta R(\mu, \mu)$ is shown in Fig. 7.18 for the QCD background and some benchmark signal models, using events that pass the baseline selections. It can be seen that there are a large number of muon SVs with large $\Delta R(\mu, \mu)$ in the QCD background compared to the signal models. Applying the selection therefore improves the signal to background ratio.

A material veto is applied to remove dimuon vertices in displacement regions where there are contributions from material interactions with the tracker. These vertices originate from nuclear interactions of hadrons with the tracker material, so they are called material vertices. Due to the geometry of the pixel tracker modules, the material vertices are displaced. Figure 7.19 (top) shows the 2-dimensional distribution of the positions of muon secondary vertices in the $x - y$ plane in a subset of data without applying material veto. It can be seen that the distribution of material vertices map out the geometry of the tracker modules.

As the material vertices are not of interest, a veto is applied to remove them. A first

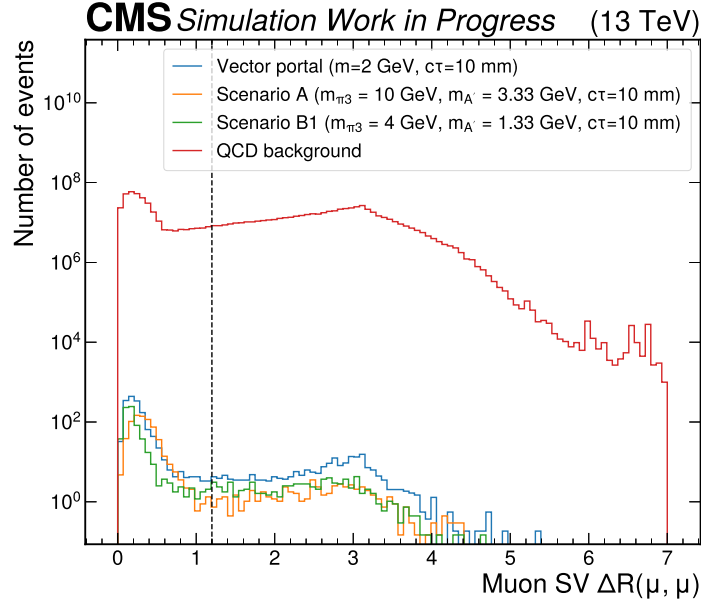


Figure 7.18: Distribution of $\Delta R(\mu, \mu)$ for the QCD background and some benchmark signal models. All events passing the baseline selections are used.

attempt of material veto was performed by applying selections on the minimum distances of the vertices from the tracker modules to remove vertices that are close to the modules. However it did not work well as some residual material effects were seen. Therefore the material veto is performed with the more robust procedure of removing all vertices in the tracker material region, i.e. removing all vertices in the displacement regions detailed below:

- $l_{xy} \in [6.4, 7.3]$ cm
- $l_{xy} \in [10.5, 11.5]$ cm
- $l_{xy} \in [15.6, 16.6]$ cm
- $27 < |z| < 52$ cm,

where l_{xy} is the transverse displacement of the secondary vertex. The regions of material veto are also illustrated in Fig. 7.20. There is only a small change in acceptance (percent level) after the material veto, which is found to have a small impact on the sensitivity. Therefore the more robust procedure is adopted to eliminate the possibility of misidentifying material vertices as signal vertices. The distribution of the positions of the muon secondary vertices in the transverse plane after the material veto is applied is shown in Fig. 7.19 (bottom).

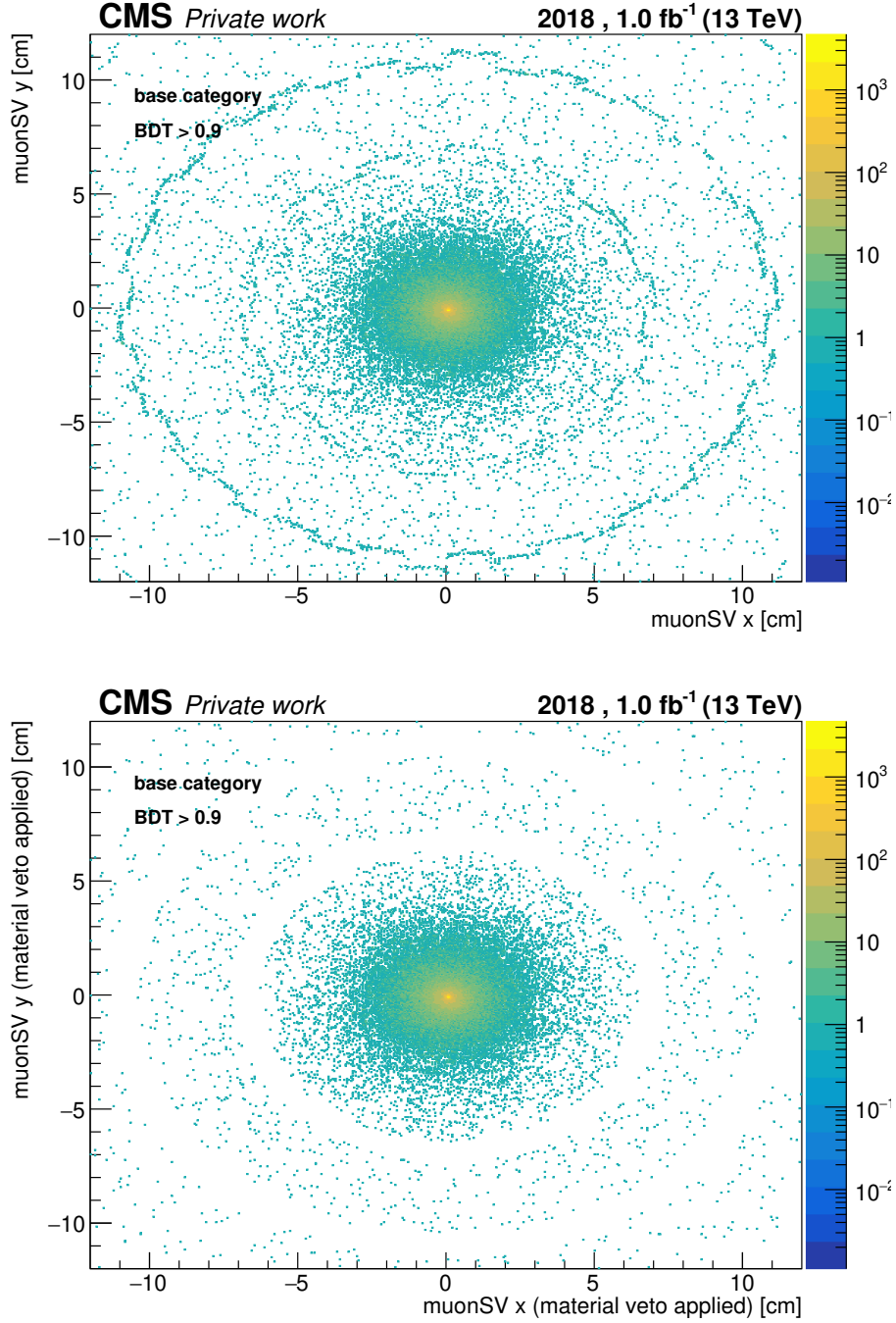


Figure 7.19: Distributions of the positions of muon secondary vertices in the $x - y$ plane for a subset of data events which pass a loosened BDT selection, before (top) and after (bottom) the material veto is applied.

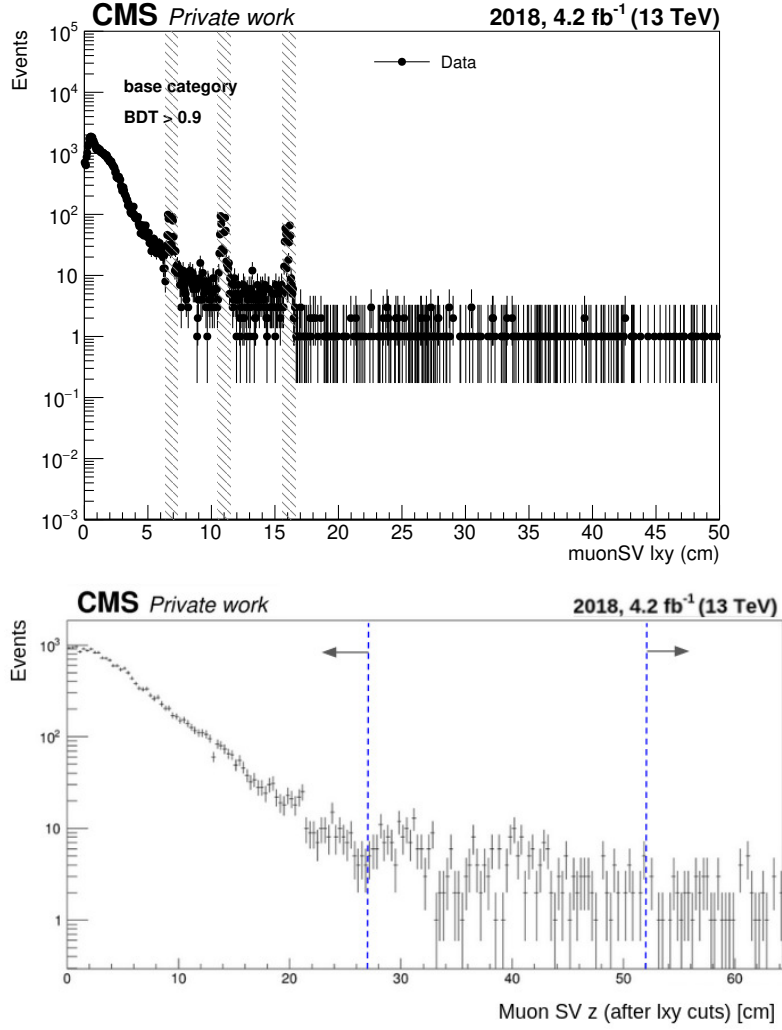


Figure 7.20: Distributions of muon secondary vertex transverse displacement (top) and z position (bottom) for a subset of data. The shaded bands in the top plot indicate the regions in transverse displacement (l_{xy}) where material veto is applied, and the bottom plot shows the selections applied to the z position of the muon secondary vertex after the l_{xy} selections are applied.

7.4.1 Summary

The event selections applied in the analysis are summarised as follow.

Baseline selections:

- B-parking displaced muon triggers
- At least one loose muon matched to the trigger muon, with
 - $|\eta| < 1.5$
 - $p_T > 9.0 \text{ GeV}$

- $\text{IP}_{\text{sig}} > 6$

Further selections:

- At least one muon secondary vertex with
 - $\chi^2 < 10$
 - $\Delta R(\mu, \mu) < 1.2$
 - one muon being trigger matched
 - both muons passing loose ID
- Material veto for muon secondary vertices in displacement regions of
 - $l_{xy} \in [6.4, 7.3] \text{ cm}, [10.5, 11.5] \text{ cm}, [15.6, 16.6] \text{ cm}$
 - $27 < |z| < 52 \text{ cm}$
- BDT selection (more details in Section 8.5)

7.5 Event categorisation

After applying event selection, event categorisation is performed to further increase the sensitivity. One of the main features of the dark shower signal models is the high multiplicity of displaced vertices, as evident from Fig. 7.4 (top left) which shows the number of muon secondary vertices (SVs) per event in benchmark signal models and the QCD background. The search targets the displaced decays of dark hadrons into pairs of muons (which is the case for the vector portal), or the sequential decays of dark hadrons into dark photons, and then into muons (which is the case for scenario A and scenario B1). The dark hadrons are produced through the hadronization of dark showers, which are initiated by the pair production of dark partons from the Standard Model Higgs boson. Therefore, at least two SVs would be expected in our signal models, and the displaced decays would result in muons with a high branching ratio. For scenario A and scenario B1, the dark photons are produced in pairs from the decays of dark pions (as shown in the Feynman diagrams in Fig. 3.3), so they are even more likely to give two dimuon vertices. On the contrary, the Standard Model background is dominated by QCD processes which do not necessarily lead to displaced decays. For example, the displaced decay of B meson in the QCD background can give the J/ψ resonance, but the probability of having two such decays in an event is expected to be small.

To increase the search sensitivity to the signal, events are first categorised based on whether there are at least two muon SVs with similar mass. This aims at targeting the

signal events where there are multiple displaced resonant decays of the dark hadrons or the dark photons. An event is classified as a multi vertex event if there is at least a pair of muon SVs with mass within 3% of each other, which is about three times of the mass resolution of the signal. Otherwise the event is classified as a single vertex event. The dimuon mass distribution for each category is shown in Fig. 7.21 for the QCD background and a benchmark signal model. The distributions use events that pass all selections, including the BDT selection. As expected, there is a much smaller background in the multi vertex category compared to the single vertex category, with a reduction in background by about three orders of magnitude. The main background in the multi vertex category is found to be the displaced decay of the J/ψ meson into muons. The number of signal events in the multi vertex category is about 20% of that in the single vertex category, and with the highly reduced background the multi vertex category has strong sensitivity.

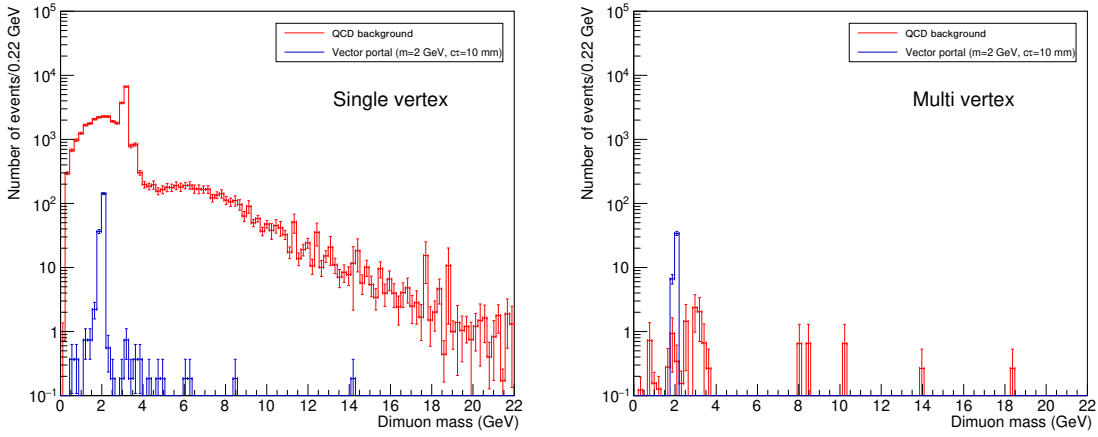


Figure 7.21: Categorising events into single vertex events (left) and multi vertex events (right) for events that pass all selections. The distribution for the QCD background uses the shape of the distribution from a looser BDT selection. For multi vertex events, the SV with the lowest χ^2 is chosen among the pairs of SVs of mass within 3% of each other.

To maximize our search sensitivity, the single vertex and multi vertex events are further categorised by the transverse displacement (l_{xy}) and pointing angle of the muon SV with the minimum χ^2 . At increasing displacement from the primary vertex, the QCD background processes are expected to diminish. The transverse displacement is used for categorisation in order to increase the sensitivity for signal models with different, longer lifetimes. Categorising in transverse displacement is found to be more beneficial to the sensitivity than categorising in transverse displacement significance of the muon SV, by studying the s/\sqrt{b} in each variable. The l_{xy} categories are defined by considering the geometry of the detector and the signal acceptance. The l_{xy} categories are defined to be $[0.0, 1.0]$ cm, $(1.0, 10.0]$ cm and beyond 10 cm. Distribution of the muon SV transverse displacement is shown in Fig. 7.22 (top) for the QCD background and some benchmark signal models. The l_{xy} distributions for benchmark signal models with different lifetimes

are shown in Fig. 7.22 (bottom), which show that the different lifetime models fit nicely into the three categories. Even though the pixel tracker extends to a transverse displacement of 16 cm, it is noted that the signal acceptance drops quickly after about 15 cm. Therefore no further categories are defined for displacement beyond that. Appendix D shows the trigger efficiency against l_{xy} for different benchmark signal models, which also demonstrates a rapid fall in the efficiency after about 15 cm.

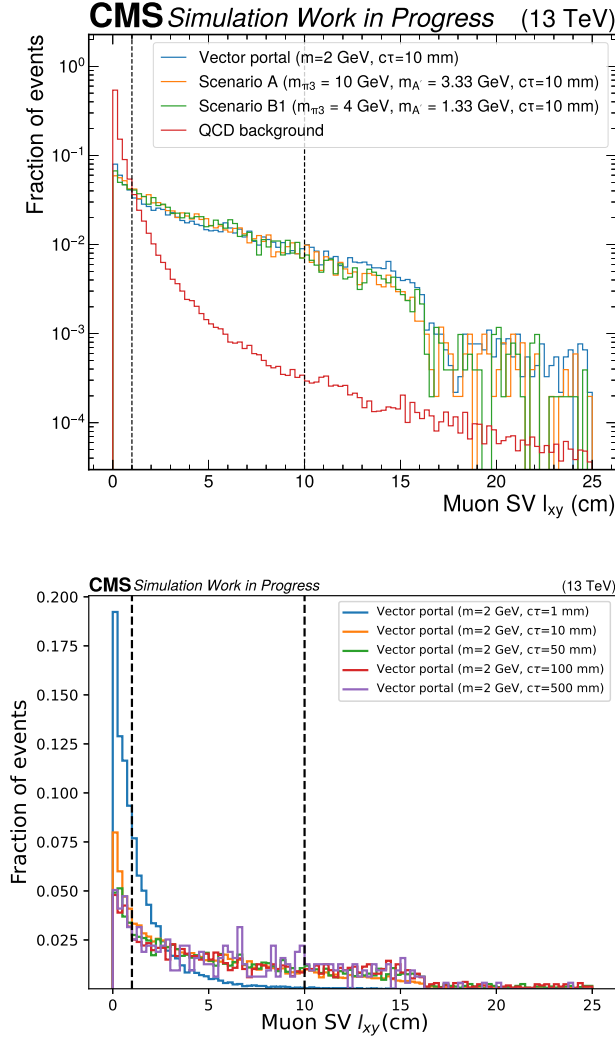


Figure 7.22: Distribution of muon SV transverse displacement for the QCD background and benchmark signal models (top), and for selected signal models of different lifetimes (bottom). All events passing the baseline selections are displayed.

In each l_{xy} bin, events are further divided into bins of pointing angle. The pointing angle of the muon SV is defined as the angle between the momentum vector of the dimuon system and the displacement vector from the primary vertex to the secondary vertex, as shown in Fig. 7.23. The pointing angle is used for categorisation to increase sensitivity to the signal models that are more pointing than the QCD background. Fig. 7.24 shows the muon SV pointing angle distribution for some benchmark signal models and the QCD

background. It can be seen that the vector portal and scenario A give very pointing muon secondary vertices, as expected. The tails of the distributions mainly come from fake vertices, which originate from particles that are misreconstructed as muons. For scenario B1, the pointing angle distribution does not peak as strongly at zero, as it is a non-pointing scenario. The pointing angle categories are chosen to be $[0.0, 0.2]$ and $(0.2, \pi]$ so as to maintain sensitivity to non-pointing scenarios as well as pointing scenarios.

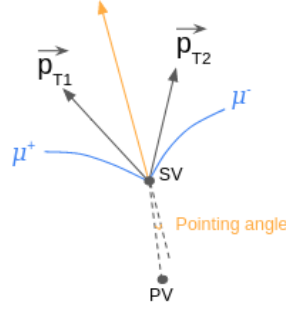


Figure 7.23: Illustration of the pointing angle of the muon SV, which is defined as the angle between the momentum vector of the dimuon system and the displacement vector from the primary vertex to the secondary vertex.

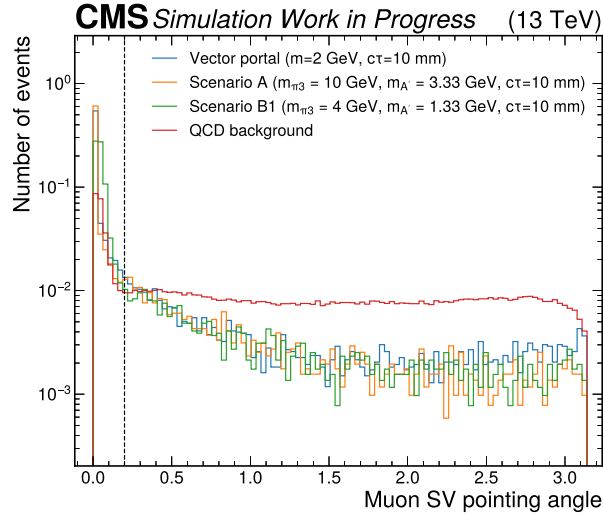


Figure 7.24: Distribution of muon SV pointing angle for the QCD background and benchmark signal models. All events passing the baseline selections are displayed.

7.5.1 Summary

The event categorisation is summarised in Table 7.7. Events are first categorised into the single vertex and the multi vertex categories, which are further split into bins of l_{xy} and pointing angle. In total there are 12 analysis categories.

Single vertex	Multi vertex
(i) $l_{xy} < 1$ cm, pointing angle < 0.2	(i) $l_{xy} < 1$ cm, pointing angle < 0.2
(ii) $l_{xy} < 1$ cm, pointing angle > 0.2	(ii) $l_{xy} < 1$ cm, pointing angle > 0.2
(iii) $1 \text{ cm} < l_{xy} < 10$ cm, pointing angle < 0.2	(iii) $1 \text{ cm} < l_{xy} < 10$ cm, pointing angle < 0.2
(iv) $1 \text{ cm} < l_{xy} < 10$ cm, pointing angle > 0.2	(iv) $1 \text{ cm} < l_{xy} < 10$ cm, pointing angle > 0.2
(v) $l_{xy} > 10$ cm, pointing angle < 0.2	(v) $l_{xy} > 10$ cm, pointing angle < 0.2
(vi) $l_{xy} > 10$ cm, pointing angle > 0.2	(vi) $l_{xy} > 10$ cm, pointing angle > 0.2

Table 7.7: Full event categorisation for the analysis. Events are first divided into the single vertex and multi vertex categories. Each of the categories is then subdivided into 6 categories based on the transverse displacement and pointing angle of the muon SV with the minimum χ^2 , resulting in 12 categories in total.

Chapter 8

Dimuon mass fit, systematic uncertainties, and results

8.1 Introduction

A fit is performed on the dimuon invariant mass in the analysis to extract the signal. Before fully unblinding the data, studies are performed to check the agreement between data and Monte Carlo to ensure good modelling in the Monte Carlo variables used for the BDT training and categorisation. As a BDT is used for distinguishing signal from background, checks are performed in a data control region to ensure that there is no sculpting effect of the BDT selection on the dimuon mass distribution.

A parametric fit is used in the analysis, where the signal and the background are described by parametric functions. Goodness of fit tests are performed to check that the signal and background shapes are fitted well by the functions. Before unblinding, the BDT selection threshold is optimised to maximise the sensitivity.

Studies on the systematic uncertainties are performed. Most of the uncertainties originate from the signal modelling, including uncertainties from the BDT selection, displaced muon identification and the trigger selection.

Using the parametric fit and taking into account the systematic uncertainties, upper limits are imposed at 95% confidence level on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ for the vector portal, scenario A and scenario B1 models.

In this Chapter, data studies are first presented, followed by signal parametrisation and background estimation. Systematic uncertainties are then discussed. Finally, the branching ratio limit results are discussed.

8.1.1 Fitting technique

A binned likelihood function is used to perform the fit, which is given by:

$$\mathcal{L}(\mu, \vec{\theta}) = \left[\prod_i \frac{(\mu s_i(\vec{\theta}) + b_i(\vec{\theta}))^{n_i}}{n_i!} e^{-(\mu s_i(\vec{\theta}) + b_i(\vec{\theta}))} \right] \cdot C(\vec{\theta}), \quad (8.1)$$

where μ denotes the signal strength (which is the parameter of interest), $s_i(\vec{\theta})$ is the signal estimate in bin i , $b_i(\vec{\theta})$ is the background estimate in bin i , and n_i is the observed data in bin i . The signal and the background estimates depend on the nuisance parameters $\vec{\theta}$, which are constrained by $C(\vec{\theta})$ in the likelihood. The signal and background rates are affected by log-normal distributions of the nuisance parameters which cannot go negative. The constraint term $C(\vec{\theta})$ is a product of normal distributions. The fit minimises the negative log likelihood (or maximises the likelihood) by using a profiled likelihood ratio test statistic

$$q_\mu = -2 \ln \frac{\mathcal{L}(\mu, \vec{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})}, \quad (8.2)$$

where $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the signal strength and the nuisance parameters that give the maximum likelihood. $\vec{\theta}(\mu)$ represents the nuisance parameters that maximise the likelihood function for a given μ (i.e. the nuisance parameters are profiled away).

The CL_s technique is used to impose upper limits on the signal strength [129, 130]. The CL_s criterion for deriving exclusion region is defined by:

$$CL_s = \frac{CL_{s+b}}{CL_b} < \alpha, \quad (8.3)$$

where

$$CL_{s+b} = \int_{q_\mu(x_{obs})}^{\infty} P(q_\mu | \mu' = \mu) dq_\mu, \quad (8.4)$$

$$CL_b = \int_{q_\mu(x_{obs})}^{\infty} P(q_\mu | \mu' = 0) dq_\mu. \quad (8.5)$$

P denotes the probability density function of q_μ , while α is the significance level of the test, which is the probability of falsely rejecting the null hypothesis given that it is true. In the analysis, 95% confidence level upper limits are imposed, which correspond to $\alpha = 0.05$.

The asymptotic approximation [131] was used at first to compute the upper limits. The approximation is based on results from Wald [132] and Wilks [133]. Wald showed that the profiled likelihood ratio test statistic q_μ is approximately parabolic for a large sample

size, while Wilks demonstrated that q_μ follows a non-central chi-square distribution for one degree of freedom.

The upper limits are obtained with toy experiments in the final results, where toy datasets are generated to create a distribution of the test statistic q_μ . The CL_s technique is then applied to obtain the limits using the probability distribution of the test statistic obtained with toys.

8.2 Data studies

8.2.1 Data and Monte-Carlo comparison

A small fraction ($\sim 2\%$) of the 2018 B-parking dataset is first unblinded to study the agreement between data and the QCD Monte Carlo (MC) simulation. About 1 fb^{-1} of data is used for the study so that the dataset is not sensitive to new physics.

The BDT output score distributions for data and the QCD MC are first compared, as shown in Fig. 8.1 (left). There is a good agreement between the shapes of the BDT score distributions for data and MC. This shows that the BDT is behaving as expected when it is deployed on data even though it is trained with simulations. A systematic uncertainty will be applied in the analysis to account for the small difference in the BDT selection efficiency in data and MC, which will be discussed in Section 8.6.2.

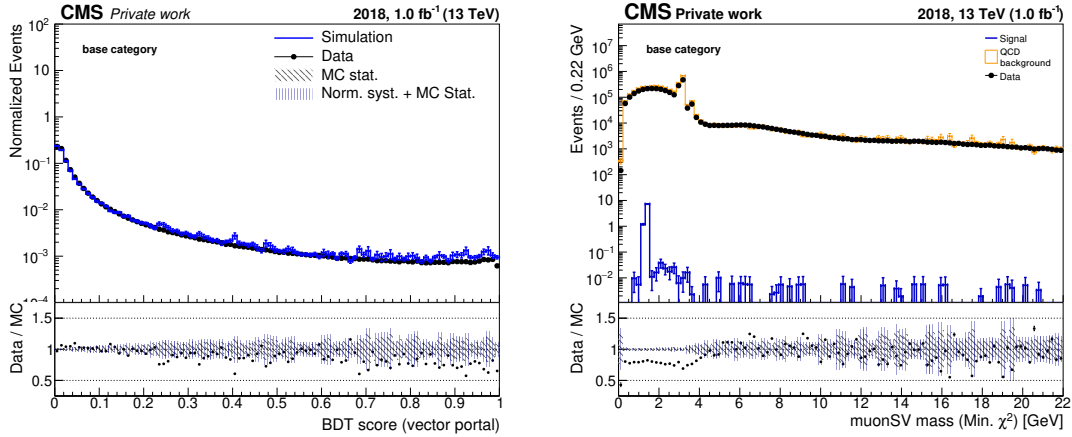


Figure 8.1: Distributions of the BDT output score (left) and the dimuon invariant mass (right) for a subset of data and the QCD background Monte Carlo. The BDT used in the left plot is trained for the vector portal signal model. Events shown pass all selections except the BDT selection and the material veto. The "base category" label means all of the twelve analysis categories combined.

The dimuon invariant mass distributions in data and MC are then studied, as shown in Fig. 8.1 (right), with a representative signal model overlaid. To further study the agree-

ment between data and MC in the dimuon mass distribution, they are split into the twelve analysis categories. The distributions are shown in Fig. 8.2 for the single vertex categories and in Fig. 8.3 for the multi vertex categories. It is noted that the agreement between data and MC is not good in the mass region below about 4 GeV. However, as the background will be estimated from data in the analysis (as will be explained in Section 8.4), the disagreement will not affect the background estimation. This also motivates the need for a data-driven background estimation method.

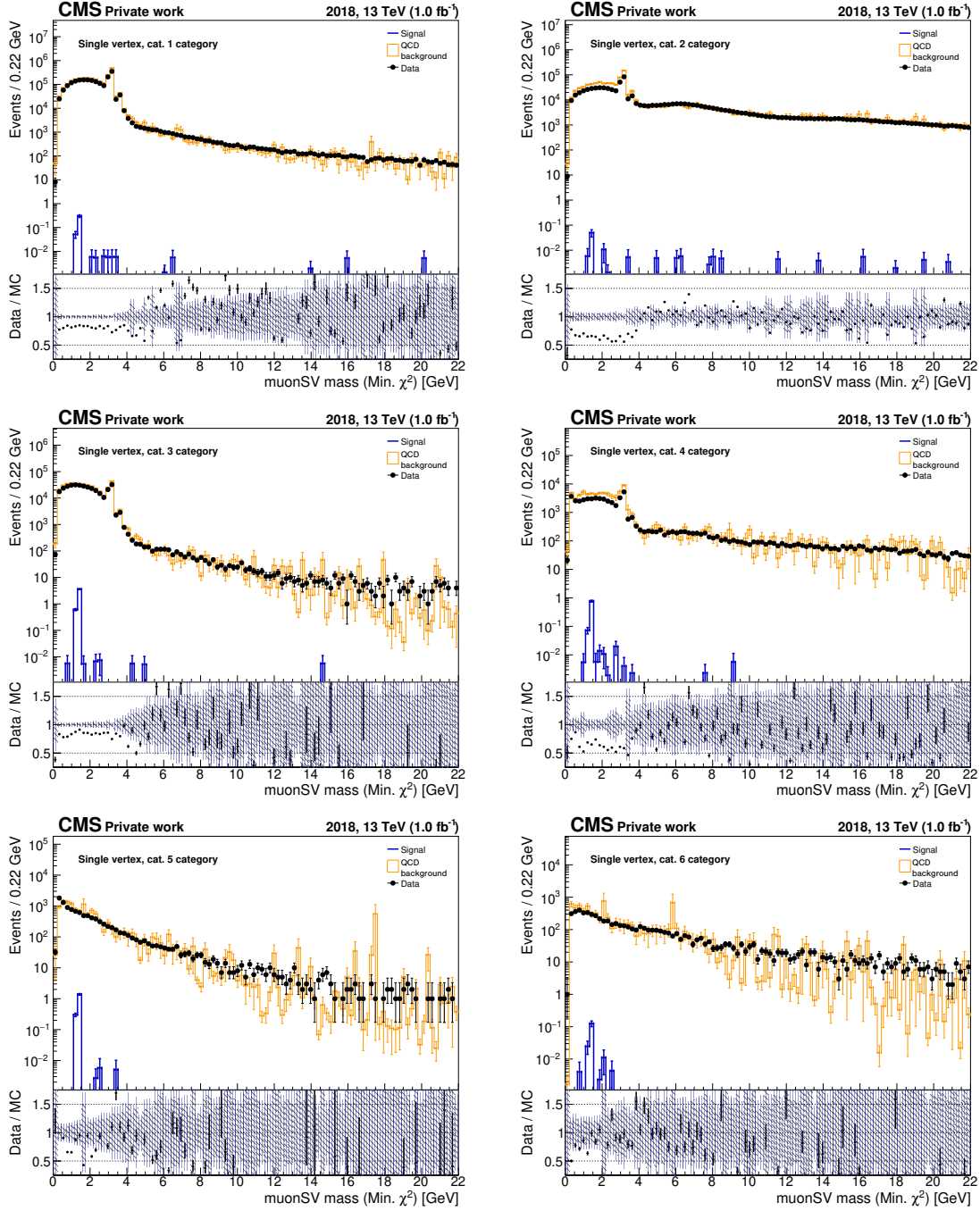


Figure 8.2: Dimuon mass distributions in the six single vertex categories for a subset of data, the QCD background Monte Carlo and a benchmark signal model from scenario A. Events shown pass all selections except the BDT selection and the material veto.

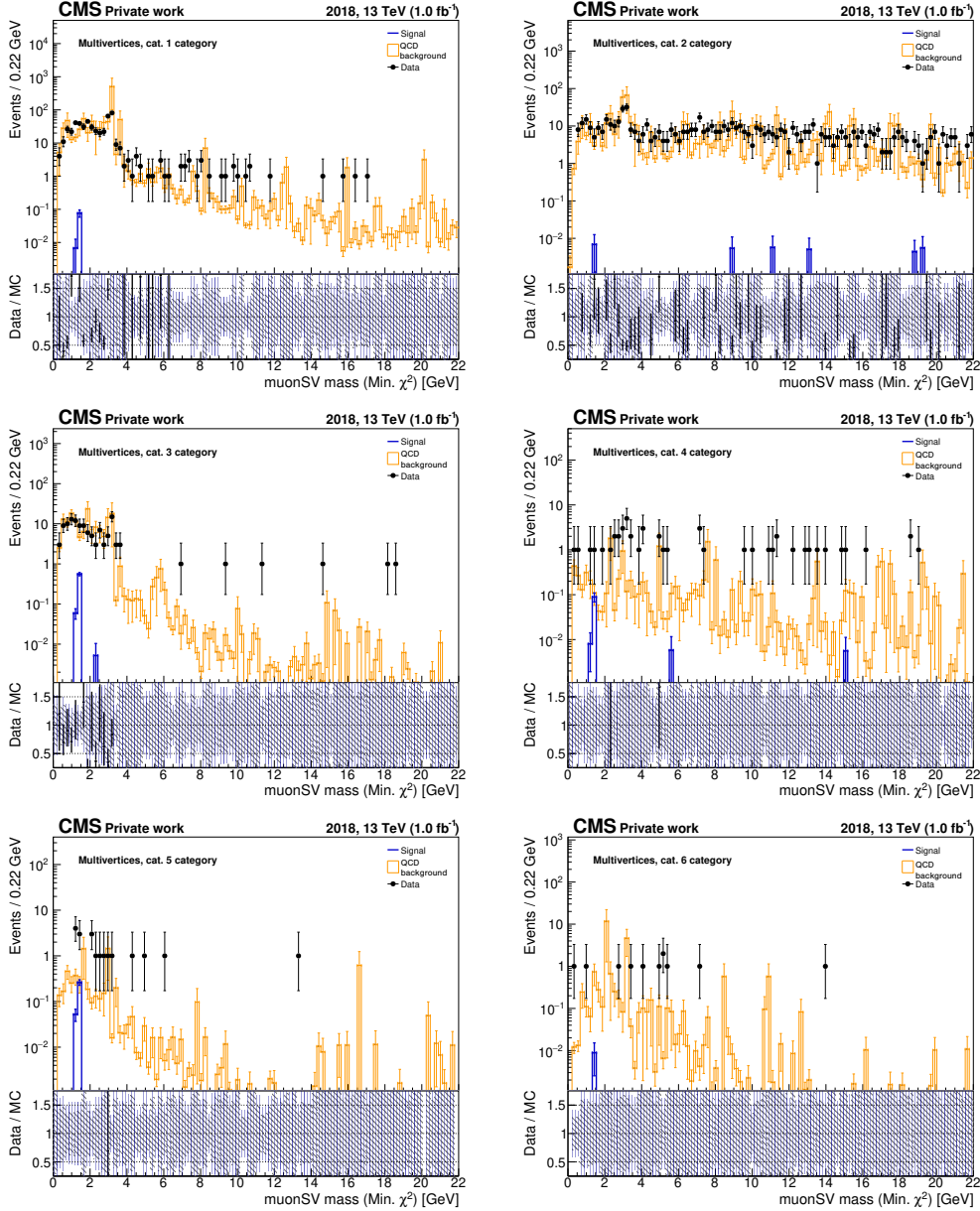


Figure 8.3: Dimuon mass distributions in the six multi vertex categories for a subset of data, the QCD background Monte Carlo and a benchmark signal model from scenario A. Events shown pass all selections except the BDT selection and the material veto.

The agreement between data and MC in variables of muon secondary vertices are also studied. The χ^2 , transverse displacement (l_{xy}), transverse displacement significance ($l_{xy} \text{sig}$) and pointing angle distributions for the muon secondary vertex (SV) of the lowest χ^2 in the event are shown in Fig. 8.4 for data and MC. There is a good agreement in general between data and MC for these variables, which are a subset of the input variables used in the BDT training, as discussed in Section 7.3.3. This shows good modelling of the variables. Moreover, the transverse displacement and pointing angle of the muon secondary vertex are used in the event categorisation. It is noted that there is some disagreement between data and MC in the $l_{xy} \text{sig}$ of the muon SV, and it further supports the choice of

using l_{xy} instead of $l_{xy} \text{sig}$ for the event categorisation. Any disagreement between data and MC in the BDT input variables will be accounted for by deriving a correction factor and an associated systematic uncertainty in the BDT selection efficiency, as will be discussed in Section 8.6.2.

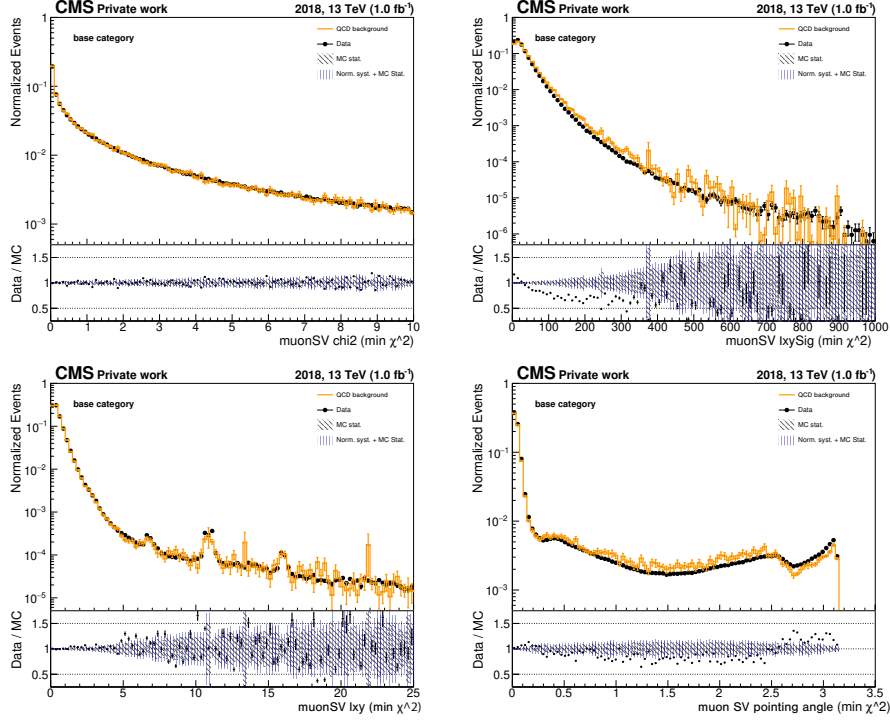


Figure 8.4: Distributions of some variables of the muon secondary vertex with the lowest χ^2 for a subset of data and the QCD background Monte Carlo. Events shown pass all selections except the BDT selection and the material veto. These plots are normalised to unity to compare the shapes of the data and MC distributions.

8.2.2 Impact of BDT selections on the dimuon mass distribution

A BDT is used in the analysis to discriminate between the signal and the background. It provides great capability in background rejection by reducing the number of background muon secondary vertices in the dimuon mass distribution. Since a fit will be performed on the dimuon mass in the analysis, it is important to check that the BDT selections do not lead to the creation of bumps in the mass distribution that would produce fake signal, i.e. there is no “sculpting” effect. A study is hence performed to investigate the effects of BDT selections on the dimuon mass distribution.

A subset of the B-parking data corresponding to 1 fb^{-1} is first used to study the dimuon distribution of opposite-sign muons. After applying the analysis preselections, effects of BDT selections on the dimuon mass distribution are investigated using mass windows of $\pm 5 \times \sigma$ of the signal mass hypothesis at $m = 2, 5, 10, 15$ and 20 GeV , where 1σ is taken to be 1% of the mass hypothesis. As will be discussed in Section 8.3, the dimuon

mass distributions of different signal hypotheses are fitted with the Voigtian function and the resolutions are found to be consistent with 1% of the signal mass. Different BDT selections (BDT > 0.95, 0.97 and 0.99) are applied to the dimuon mass distribution. The ratio of the resulting mass distribution in each case to the inclusive distribution is obtained. A fit is then performed to the ratio by using either a linear function or a constant function. The results are shown in Fig. 8.5.

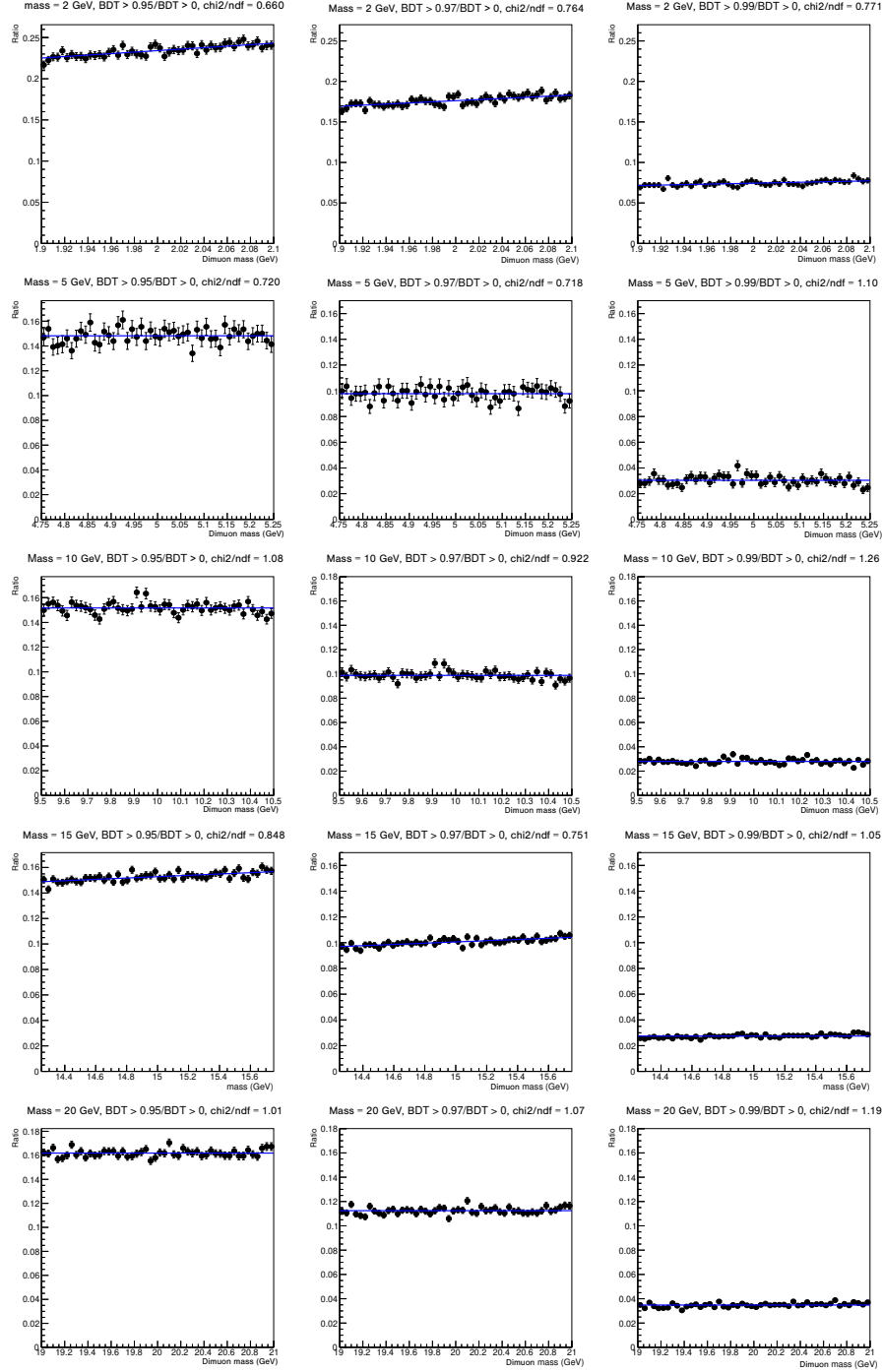


Figure 8.5: Fits of the ratio of the dimuon mass distribution with different BDT selections to the inclusive distribution for different signal mass hypotheses using a subset of data. The χ^2/ndf for each fit is also shown. No sculpting of the mass distribution is observed.

The χ^2/ndf (reduced chi-squared) is used to assess the goodness of fit. The chi-squared is defined as $\chi^2 = \sum_i (d_i - f_i)^2 / \sigma_i^2$, where d_i represents the data point, f_i represents the prediction and σ_i denotes the uncertainty of the data point. The “ndf” corresponds to the number of degrees of freedom, which is given by the number of parameters of the fit subtracted from the number of non-zero bins in the distribution. The χ^2/ndf of each fit shown in Fig. 8.5 is plotted in Fig. 8.6 against the dimuon mass. It shows that a straight line fit to the ratio of the mass distribution with the BDT selection to that without the selection is a good fit for all the BDT selections and mass windows considered, with χ^2/ndf close to 1. Therefore, no sculpting of the mass distribution by the BDT selections is observed.

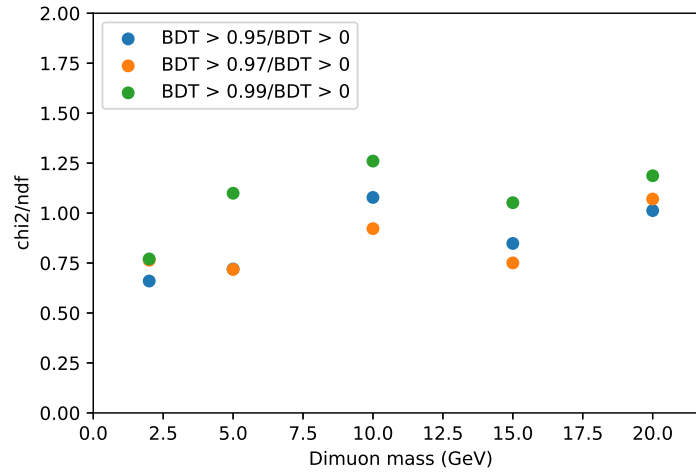


Figure 8.6: The χ^2/ndf of the fits to the ratio of the dimuon mass distribution in a subset of data with different BDT selections to that without the selections are plotted for different signal mass hypotheses.

As the study is performed before unblinding, there are a limited number of events that are available, especially with the BDT working point in the analysis which gives 10^{-4} background efficiency. To demonstrate there is no sculpting with such stringent BDT selection, a check is performed by applying the BDT selection of $\text{BDT} > 0.997$ in a data control region with same-sign muons. Similar to before, the ratio of the dimuon mass distribution with the BDT selection to the inclusive distribution is obtained inside a sliding mass window of $\pm 5 \times \sigma$ of the signal mass hypothesis across a mass range of 0.33 to 20 GeV, which is the mass range of the search. All same-sign muon pairs in each event are included in the dimuon mass histograms in order to increase the statistical precision. A fit to the ratio of the mass distribution with the BDT selection to the inclusive distribution is performed for different signal mass hypotheses by using either a linear function or a constant function. The results are shown in Fig. 8.7 and 8.8, which also show the χ^2/ndf of each fit. It can be seen that the straight line fits are good fits in general, taking into

account the statistical uncertainties of the data points, and most of them have χ^2/ndf close to 1. Therefore, no sculpting of the dimuon mass is observed in the same-sign control region using the analysis BDT selection across the mass range of the search.

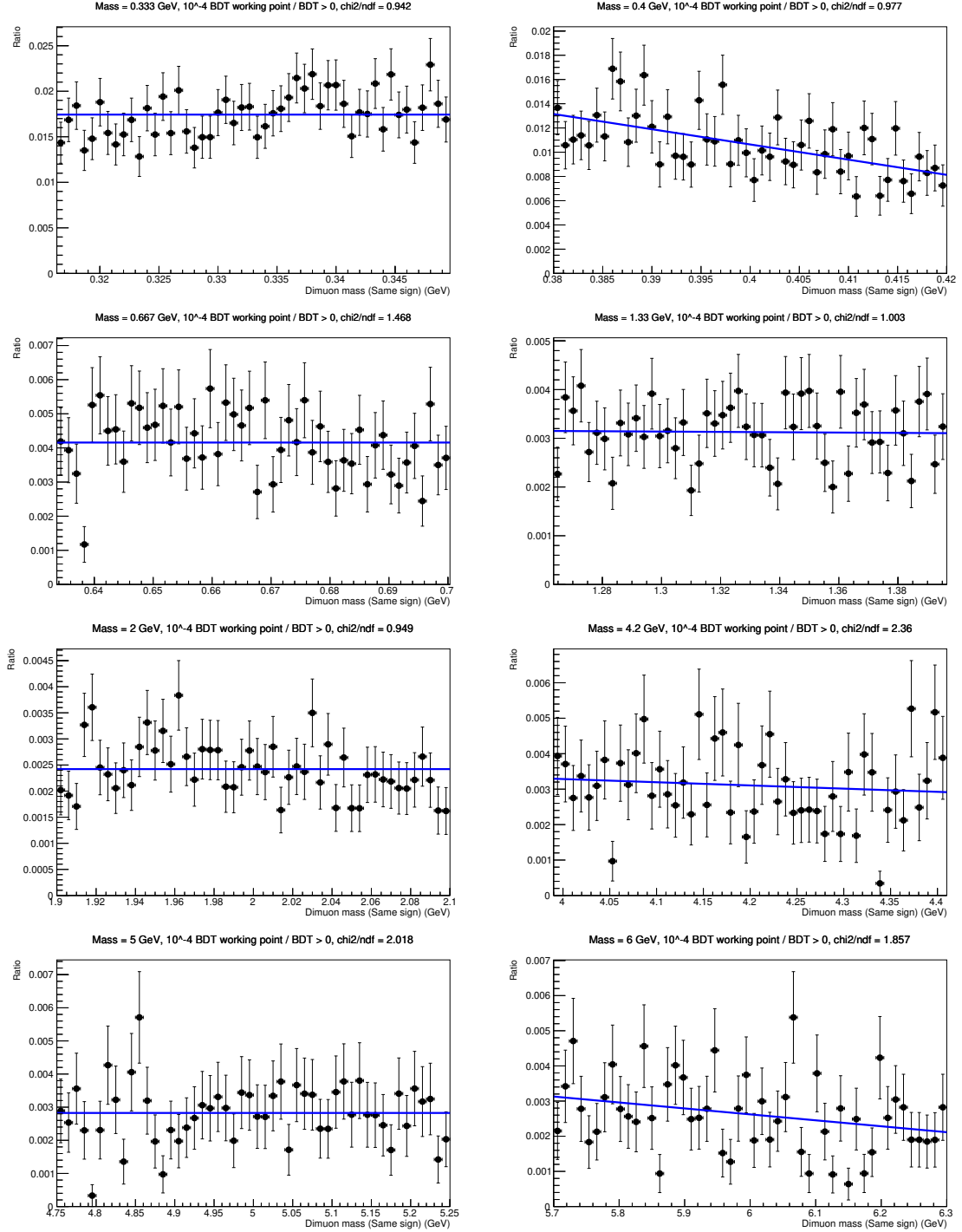


Figure 8.7: Fits of the ratio of same-sign dimuon mass distribution with the analysis BDT selection to the inclusive distribution for different signal mass hypotheses, using a subset of data. The χ^2/ndf for each fit is also shown. No sculpting of the dimuon mass is observed.

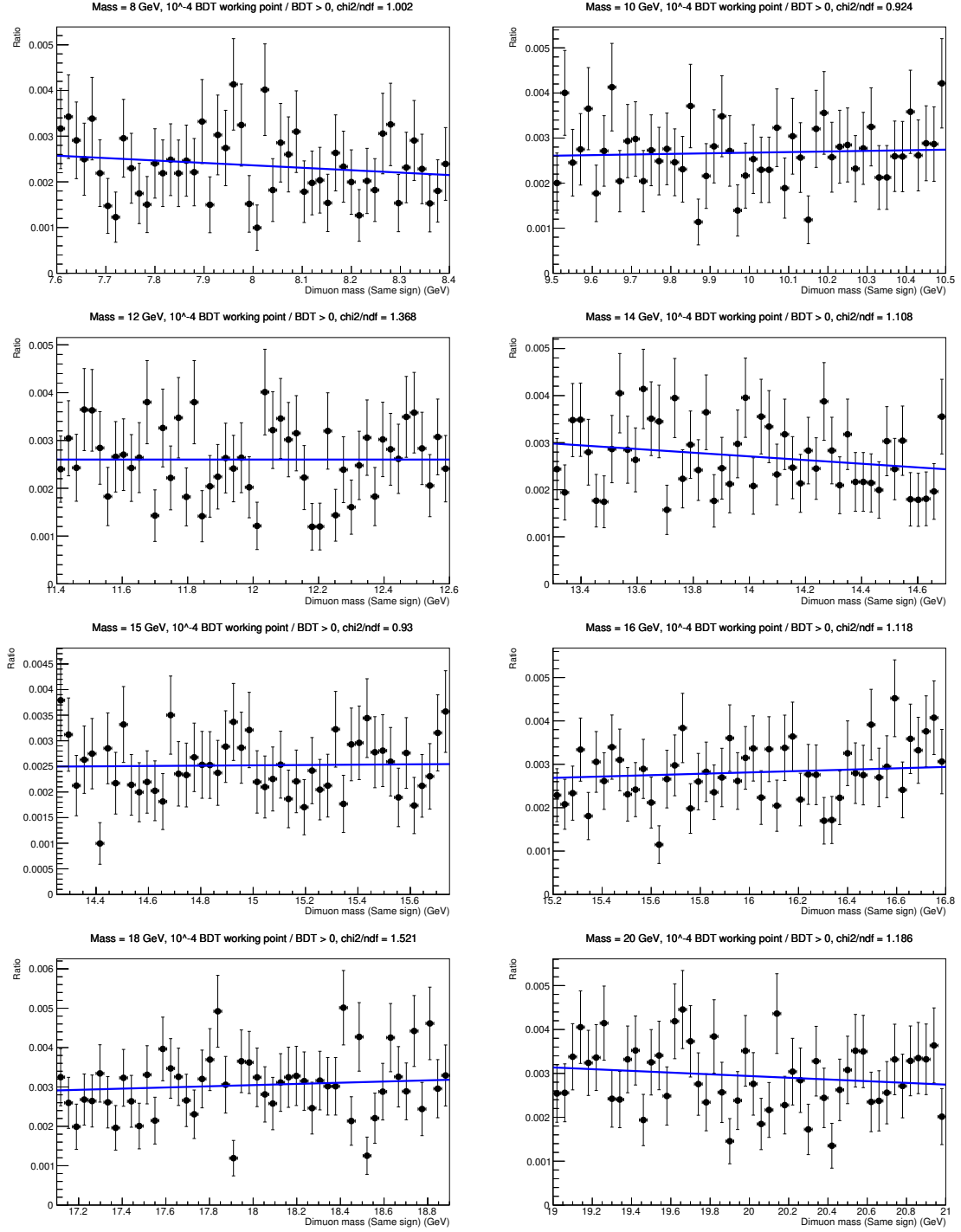


Figure 8.8: Fits of the ratio of same-sign dimuon mass distribution with the analysis BDT selection to the inclusive distribution for different signal mass hypotheses, using a subset of data. The χ^2/ndf for each fit is also shown. No sculpting of the dimuon mass is observed.

8.3 Signal parametrisation

As the dark shower signal models produce a large number of dark hadrons and dark photons which undergo displaced decays into muons, they give resonance signatures in the

dimuon invariant mass distribution. The resonance peaks are good distinguishable features of the signal compared to the combinatorial background from QCD. Therefore, the search is carried out by performing a maximum likelihood fit on the dimuon mass distribution to look for any peaking features from the signal, which is also known as a "bump hunt". To provide a more accurate characterisation of the signal and the background, parametric functions are used to model the mass distributions.

A Voigtian function (which is the convolution of a Gaussian function and a Lorentzian function) is used to fit the signal dimuon mass distribution. The Voigtian function is parametrised by the mean μ , the Gaussian width σ and the Lorentzian width γ . The fits are performed for each signal mass and lifetime hypothesis in different categories. Fits were also attempted using the Gaussian function, but it was found that using the Voigtian function better captures the sharp peak in the distribution due to its Lorentzian component and improves the goodness of fit, and hence it is chosen as the fitting function. Example fits are shown in Fig. 8.9 for each lifetime hypothesis in the vector portal with $m_{\tilde{W}} = 5$ GeV, and in Fig. 8.10-8.11 for each lifetime hypothesis in scenario A with $m_{A'} = 1.33$ GeV. As seen in Fig. 8.9-8.11, the χ^2/ndf of the fits are mostly close to one, which indicate the Voigtian function produces good fits to the signal dimuon mass distributions.

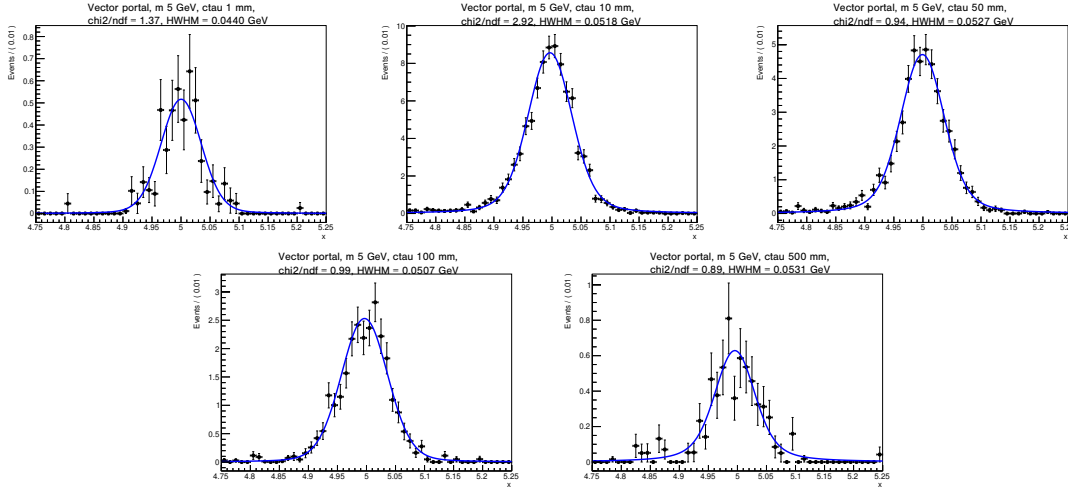


Figure 8.9: Fits of signal dimuon mass distributions for events in single vertex category 4 ($1 < l_{xy} < 10$ cm, pointing angle > 0.2). Events shown are for the vector portal models with $m_{\tilde{W}} = 5$ GeV and different lifetimes which pass all selections. The χ^2/ndf and the HWHM for each fit are also shown.

The widths of the fits are also studied. It is observed that although the σ and the γ differ among the signal fits, the half-width at half maximum (HWHM), which is a function of σ and γ , stays roughly the same. The HWHM is a measure of the width of the fit, defined as half the width of the fitted function when it is at half its maximum value. For the Voigtian function, it is given by $\text{HWHM} \simeq (0.5346f_L + \sqrt{0.2166f_L^2 + f_G^2})/2$, where $f_L = 2\gamma$ and $f_G = 2\sigma\sqrt{2\ln(2)}$ are the full-widths at half maximum of the Lorentzian function and the

Gaussian function respectively. It is found that the HWHM is about 1 % of the signal mass hypothesis, as shown in Fig. 8.9-8.11.

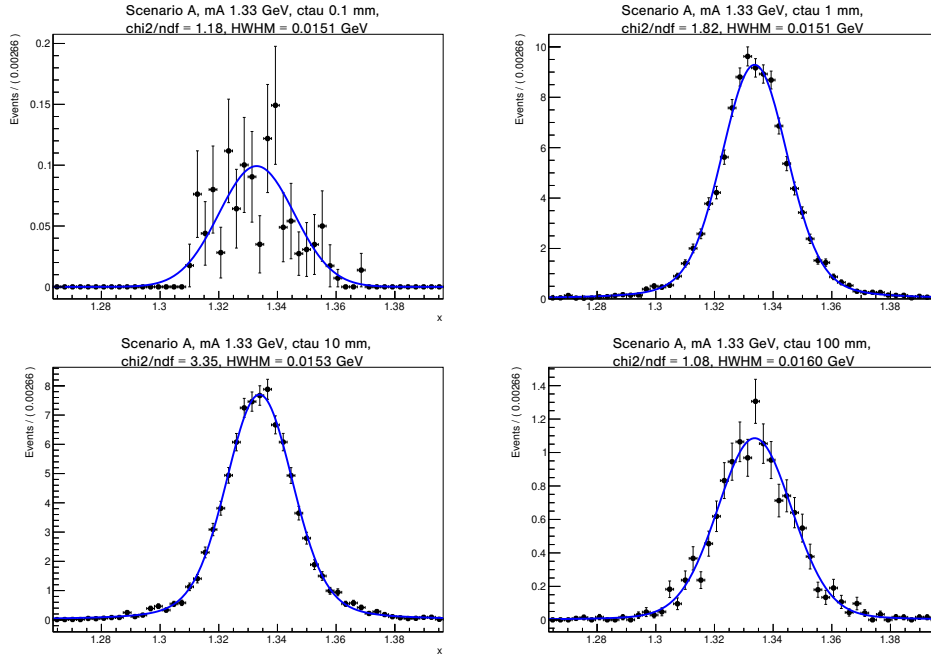


Figure 8.10: Fits of signal dimuon mass distributions for events in single vertex category 3 ($1 \text{ cm} < l_{xy} < 10 \text{ cm}$, pointing angle < 0.2). Events shown are for the scenario A models with $m_{\pi 3} = 4 \text{ GeV}$, $m_{A'} = 1.33 \text{ GeV}$ and different lifetimes which pass all selections. The χ^2/ndf and the HWHM for each fit are also shown.

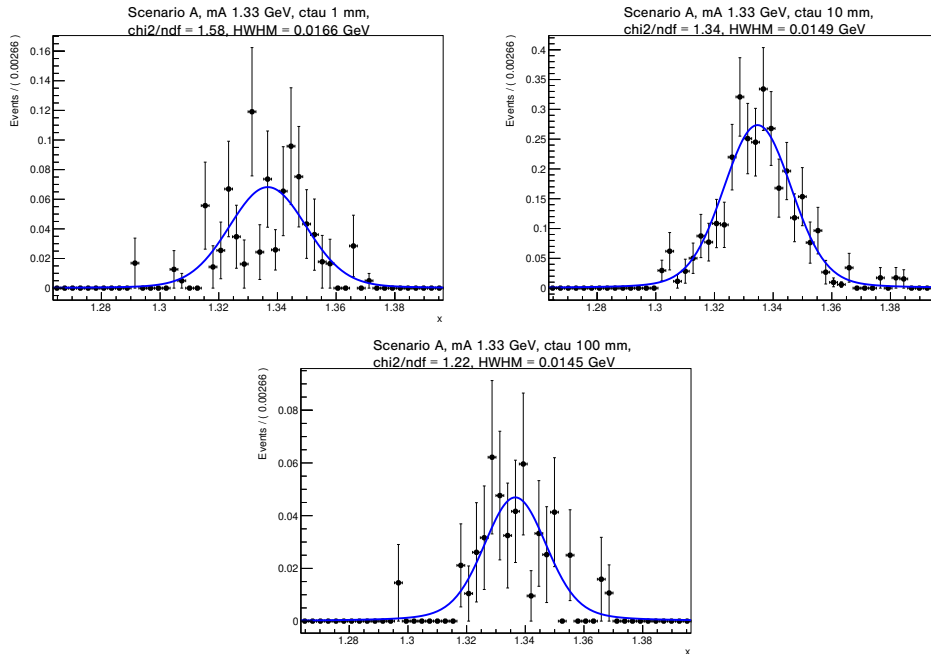


Figure 8.11: Fits of signal dimuon mass distributions for events in single vertex category 4 ($1 \text{ cm} < l_{xy} < 10 \text{ cm}$, pointing angle > 0.2). Events shown are for the scenario A models with $m_{\pi 3} = 4 \text{ GeV}$, $m_{A'} = 1.33 \text{ GeV}$ and different lifetimes which pass all selections. The χ^2/ndf and the HWHM for each fit are also shown.

To study the resonance widths for different signal mass and lifetime hypotheses, the dimuon relative resonance σ is plotted as a function of the dimuon mass for different signal models of the vector portal, scenario A and scenario B1, as shown in Fig. 8.12. The relative resonance σ is defined by the σ divided by the signal mass hypothesis, which represents the mass resolution of the signal. In the fits, the Lorentzian width γ is fixed to be $0.005 \times \text{mass}$ for the vector portal, and 0.005 for scenario A and scenario B1 to reduce the number of parameters, since it is found to have no effect on the goodness of fit (by looking at the χ^2/ndf) of the signal fit. From Fig. 8.12, it is found that the resonance width is roughly constant with small variations for different mass and lifetime models. Therefore an uncertainty will be assigned when determining the resonance widths of the different signal models.

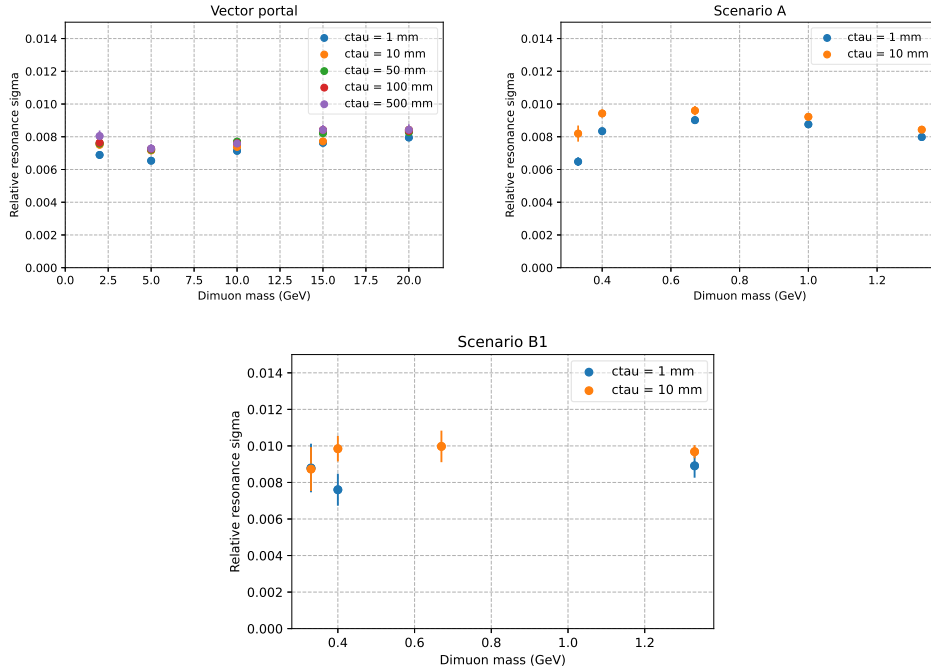


Figure 8.12: Relative resonance width σ as a function of the dimuon mass for representative lifetime models in the vector portal (upper left), scenario A (upper right) and scenario B1 (lower).

To further study the resonance widths in different event categories, the relative resonance σ is plotted as a function of the dimuon mass for representative lifetime models of the vector portal in example event categories with high statistical precision, as shown in Fig. 8.13. It is found that the variations in the resonance widths are small among the categories. The widths are consistent between the single and multi vertex categories. This is expected as the resonance width should depend on the properties of the muon secondary vertex itself instead of the vertex multiplicity of the event. Taking into account the distribution of the relative resonance σ in different categories for different signal mass and lifetime hypotheses, the relative resonance σ is found to be consistent with 0.0075 ± 0.001 for the vector

portal, and 0.0085 ± 0.0015 for scenario A and scenario B1.

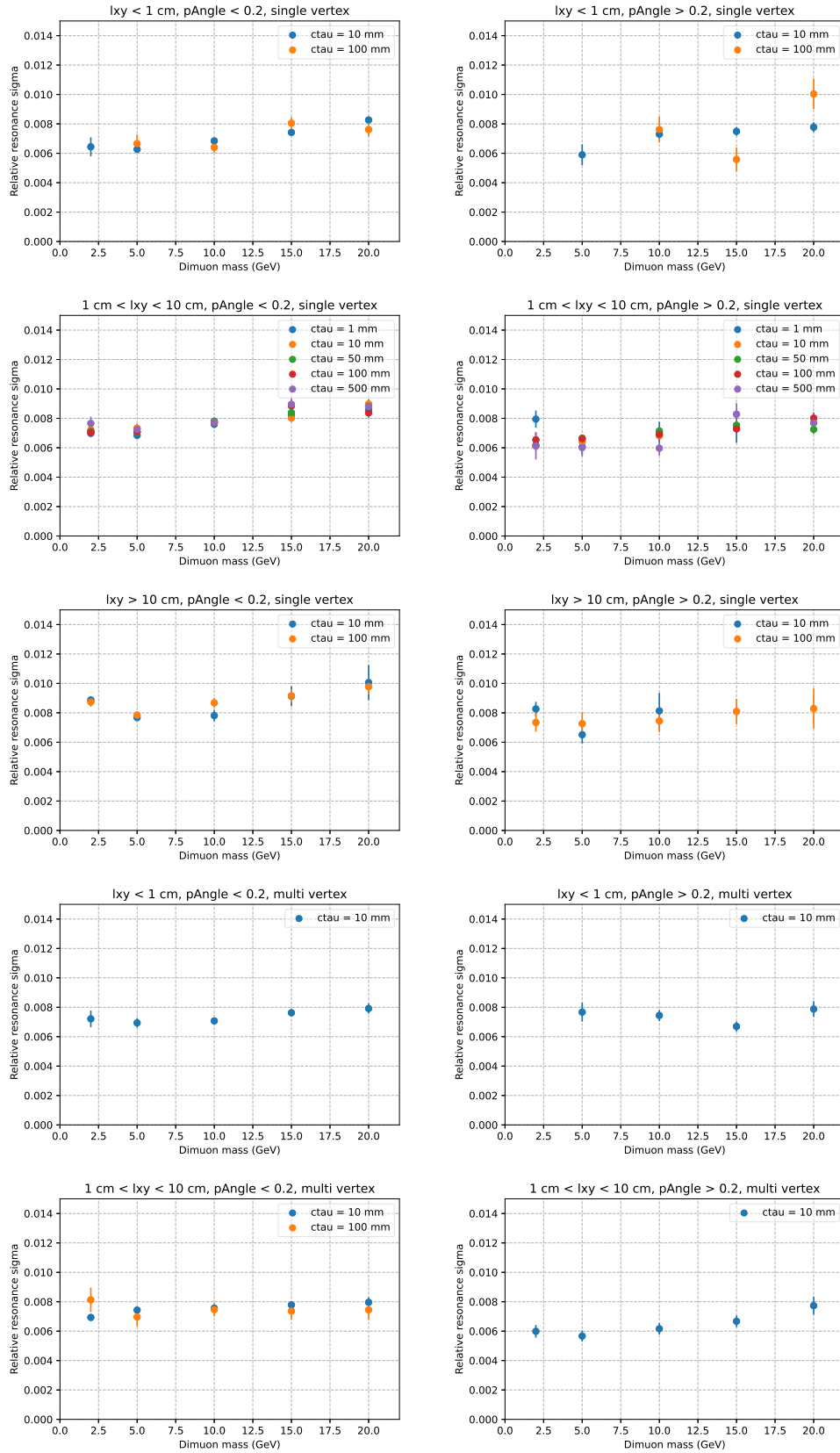


Figure 8.13: Relative resonance width σ as a function of the dimuon mass in different event categories for representative lifetime models in the vector portal.

8.4 Background estimation

In order to be sensitive to any potential signal, a sliding window of $\pm 5 \times \text{HWHM}$ of the signal resonance is used for the search. The HWHM is taken to be 1% of the mass, which is found to be consistent with the signal resolution by fitting different signal mass distributions with the Voigtian function, as discussed in Section 8.3. For the background fit, an envelope of functions are used [134], which are the first order polynomial, exponential and power law functions. The functions are found to be sufficient to fit different background shapes by performing goodness of fit tests, due to the small background in most of the event categories. As there is limited statistical precision in the background Monte Carlo simulations, the background estimation is performed using data sidebands. The signal region is defined to be $\pm 2 \times \text{HWHM}$ around the signal mass hypothesis. The background is estimated by fitting the envelope of functions to the data sidebands, which are regions in the mass window that are outside the signal region, so as to avoid any bias in the background estimation. Some example background fits are shown in Fig. 8.14. At each point in the likelihood scan, the fitting function is chosen to maximise the likelihood. Discrete profiling is used to account for the systematic uncertainty from the choice of the function [134], which will be discussed in Section 8.6.

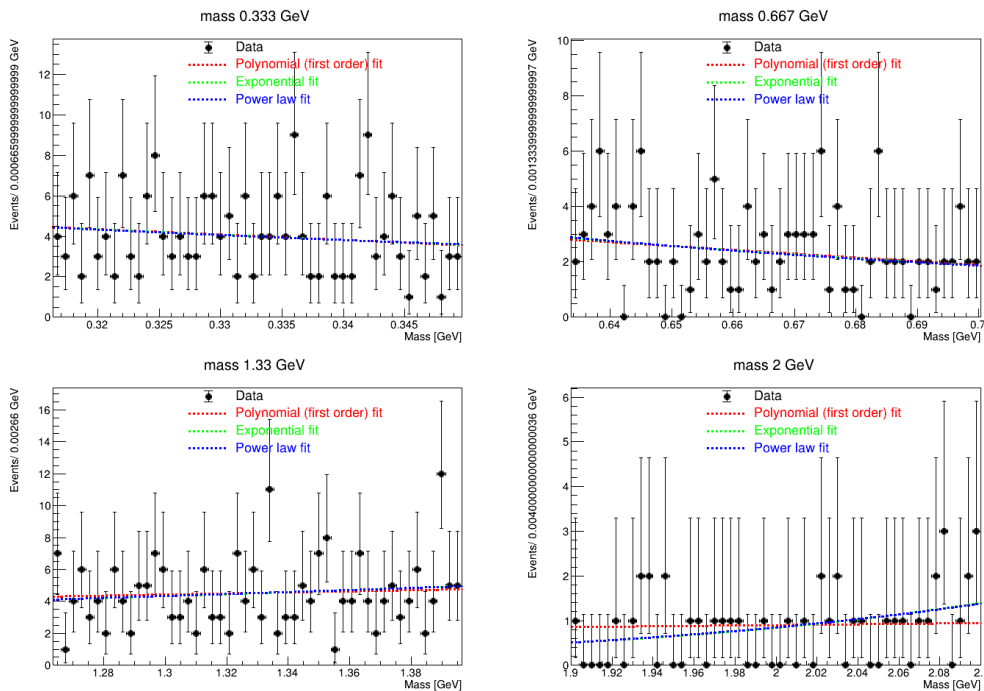


Figure 8.14: Background fits of dimuon mass distributions in data at example signal mass hypotheses in the single vertex category 3.

For mass windows with very low background, a uniform function is used instead of the envelope of functions to ensure stability of fit. Following a similar fitting strategy used in previous dimuon searches [135, 136], a uniform function is used when there are fewer

than 10 events in the data sidebands. It was found that fitting with this approach gives upper limits that agree with those obtained from a simple cut-and-count method, which uses the signal and data counts as inputs to the Poisson likelihood.

To check that the set of functions is sufficient to cover different background shapes, goodness of fit tests are performed. For example, in single vertex category 3 at $m = 1.33$ GeV, the χ^2/ndf of the fits are 0.64, 0.64 and 0.65 for the first order polynomial, exponential and power law functions respectively. Further goodness of fit tests are performed with the envelope of functions in all event categories with the saturated algorithm, which is an algorithm that uses a similar test statistic to χ^2 but it can be computed for binned data that are not Gaussian distributed [137]. p-values are computed for fits in different mass windows in each category, as shown in Fig. 8.16 and Fig. 8.17. The p-values indicate how compatible the observed data are with the probability density function (PDF) that is used for the fit. Good p-values are obtained for all fits. It is noted that some fits have p-values of 1, which correspond to mass windows with zero observed events. In such cases the model PDF is a constant function at zero, which exactly matches the data distribution.

Bias tests are also performed to study the potential bias in the choice of functional form used to parametrise the background. Toy datasets are generated using the exponential function to parametrise the background component, with an injected signal of strength $r_{\text{injected}} = 5$, where a reference signal strength of $r_{\text{injected}} = 1$ corresponds to the $+2\sigma$ threshold of the expected limits obtained with 10% of data. A fit is then performed using the full envelope of functions, and the pull is evaluated in each toy dataset, which is defined by $\frac{(r_{\text{measured}} - r_{\text{injected}})}{\sigma_r}$, where r_{measured} and σ_r are the measured signal strength and its uncertainty respectively. The bias due to the choice of functional form is taken to be the mean of the pull distribution, which is shown Fig. 8.15 for three representative signal lifetime hypotheses. Above $m = 2$ GeV, there are fewer than 10 observed events in the mass windows for each category, so a uniform function is used for the fit, as defined in our analysis procedure. Therefore only a single functional form is used for those fits, and bias tests are not informative. The results for the envelope fits show that the bias is within 5%, which is reasonable.

The dimuon mass distributions for data and two representative benchmark signal models in the vector portal are shown in Fig. 8.18 and Fig. 8.19 for all categories. As expected, very small background is observed in the multi vertex categories compared to the single vertex categories, so they provide good sensitivity. The categories with smaller pointing angle are also found to give more favourable signal to background ratio, which is due to the muon vertices in the signal models being more pointing than those in the background.

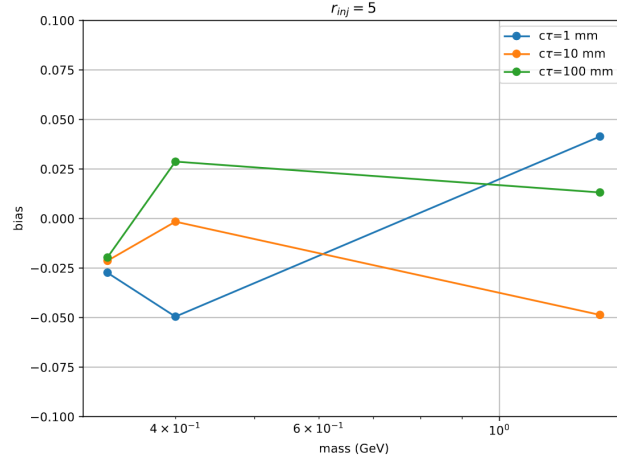


Figure 8.15: Bias values for $r_{\text{injected}} = 5$ against the dimuon mass for three representative lifetime hypotheses in scenario A.

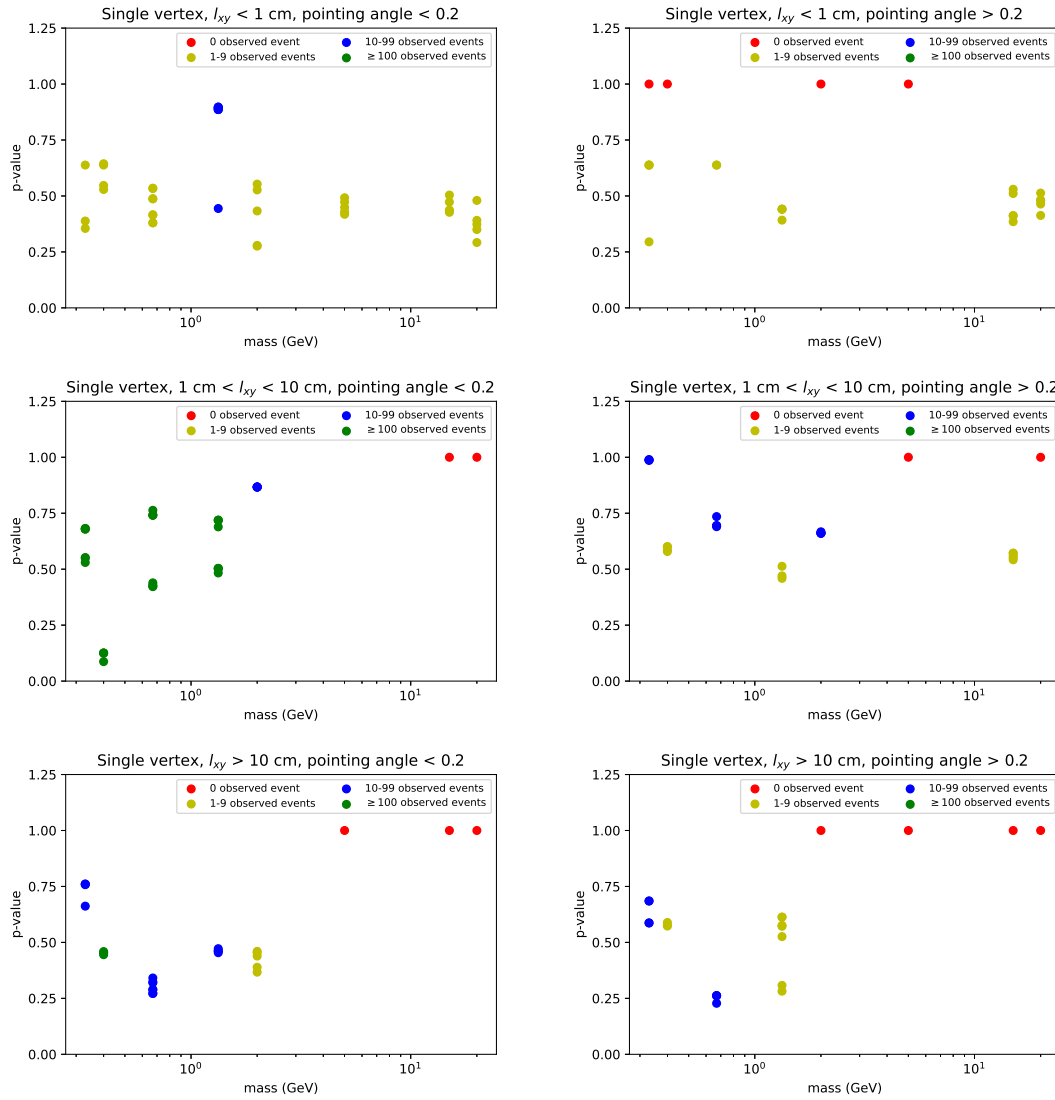


Figure 8.16: Goodness-of-fit test (using the saturated model) p-values as a function of mass in the single vertex categories. Each point is coloured red, orange, blue or green if the number of observed events in the mass window is 0, 1-9, 10-99 or above 99 respectively.

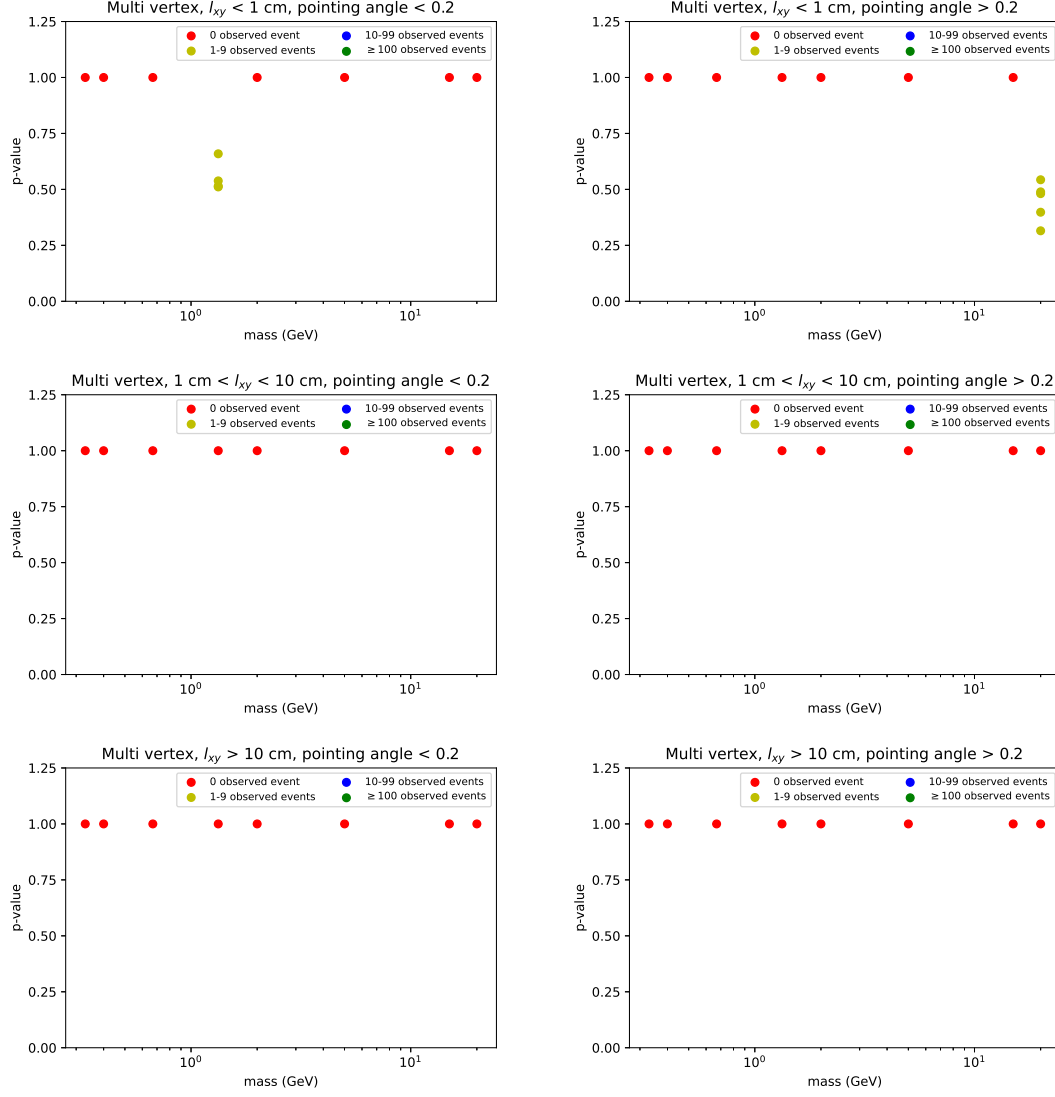


Figure 8.17: Goodness-of-fit test (using the saturated model) p-values as a function of mass in the multi vertex categories. Each point is coloured red, orange, blue or green if the number of observed events in the mass window is 0, 1-9, 10-99 or above 99 respectively.

When performing the fit to the dimuon mass, the mass window is required not to overlap with any Standard Model resonances in order to avoid identifying the known resonances as coming from the dark shower signal models. There could also potentially be fake dimuon resonances that come from pions being misidentified as muons, for example through the decay of the K_S meson. Therefore a number of “masked” regions in the dimuon mass distribution are defined in Table 8.1. The location and the size of the masked regions were determined by fitting to the resonances, and each region is defined to be $\pm 5\sigma$ around the mean resonance mass [135].

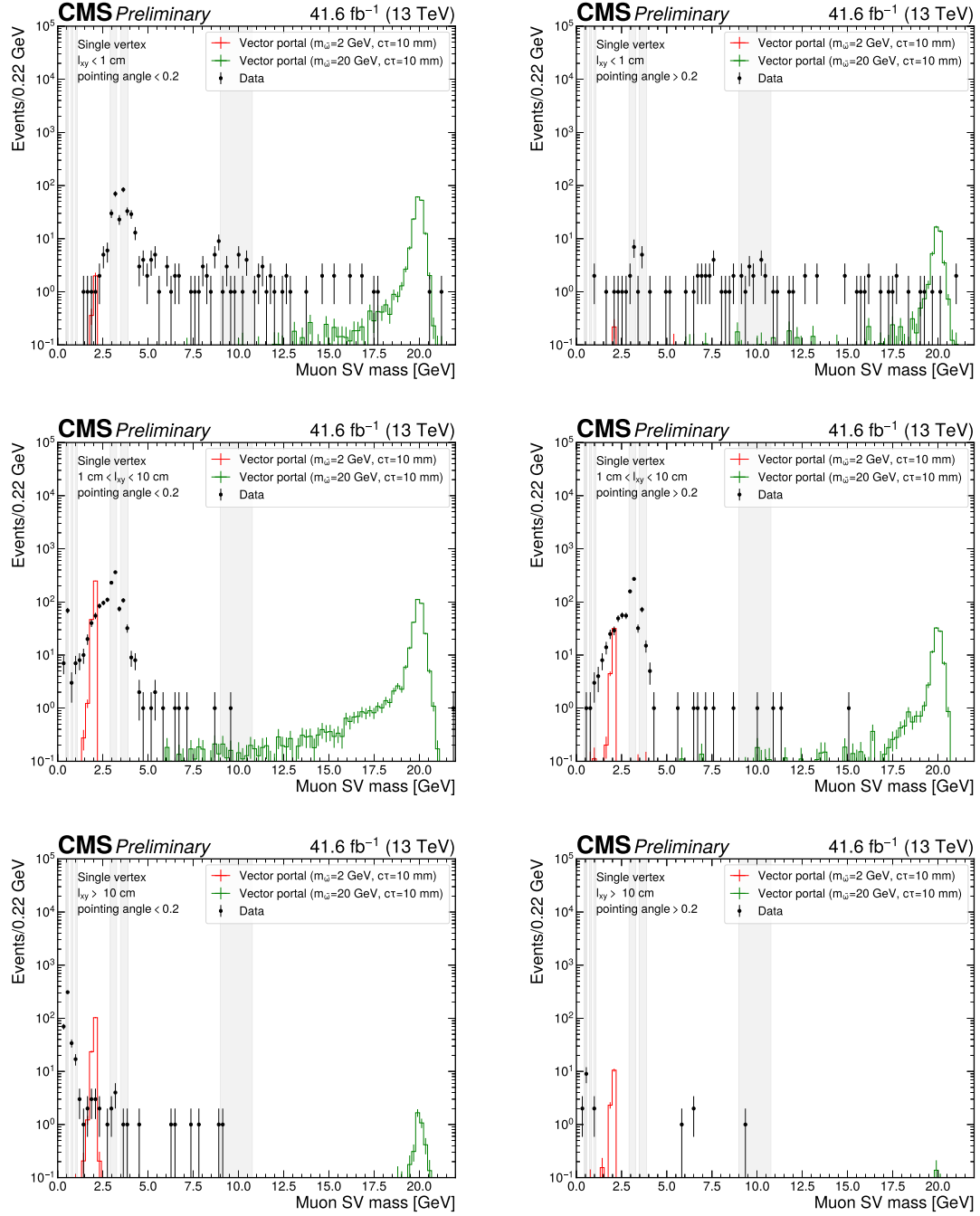


Figure 8.18: Dimuon mass distributions in the single vertex categories for data and two benchmark signal models in the vector portal, using events that pass all selections. The shaded regions indicate mass regions of known Standard Model resonances, which are masked in the search. The masked regions are detailed in Table 8.1.

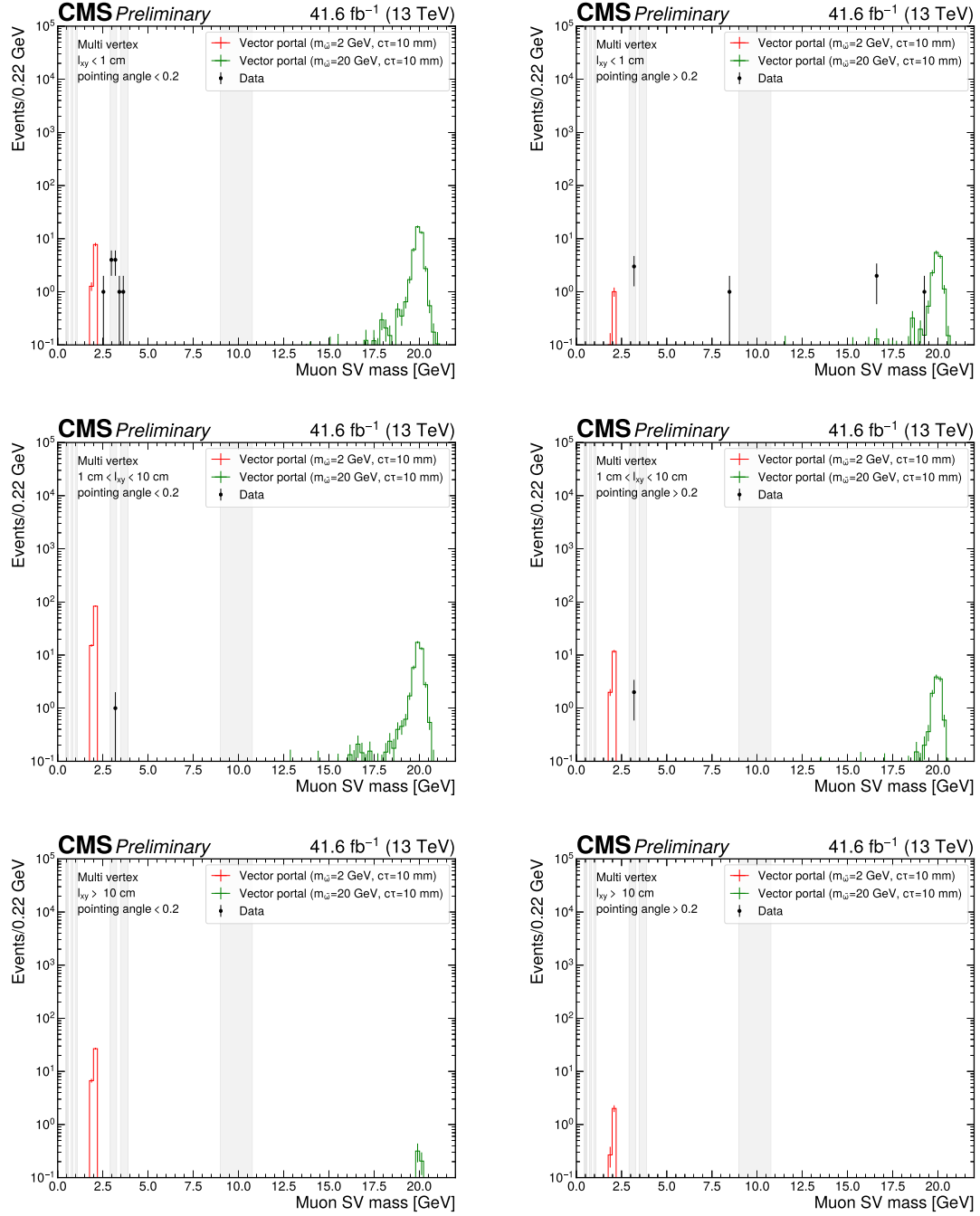


Figure 8.19: Dimuon mass distributions in the multi vertex categories for data and two benchmark signal models in the vector portal, using events that pass all selections. The shaded regions indicate mass regions of known Standard Model resonances, which are masked in the search. The masked regions are detailed in Table 8.1.

Resonance	Lower bound [GeV]	Upper bound [GeV]
K_S	0.43	0.49
η	0.52	0.58
ρ/ω	0.73	0.84
ϕ	0.96	1.08
J/ψ	2.91	3.27
$\Psi(2S)$	3.47	3.89
$\Upsilon(1S)$	8.99	9.87
$\Upsilon(1S)$	9.61	10.39
$\Upsilon(3S)$	9.87	10.77

Table 8.1: List of Standard Model resonances and the corresponding mass windows that are "masked" in the analysis.

8.5 Optimisation of the Boosted Decision Tree selection

Before unblinding, the BDT selections are optimised for the different signal models by computing the expected upper limits at 95% confidence level with different BDT thresholds. The limits are obtained by performing a maximum likelihood fit, using parametric functions to describe the signal and background shapes as discussed in Sections 8.3 and 8.4. The limits are set using the CL_s criterion. The background estimation is performed using the data sidebands, as will be implemented in the final results. The expected limit is computed for benchmark signal mass hypotheses with a sliding mass window, such that the data sidebands used in the background estimation in each case do not overlap with any of the signal mass regions considered. At each mass hypothesis, different lifetime hypotheses are studied to observe the impact of different BDT thresholds to the sensitivities of different lifetime models. The sensitivity scans are shown in Fig. 8.20, Fig. 8.21 and Fig. 8.22 for some representative signal models of the vector portal, scenario A and scenario B1 respectively, which show the expected upper limits on the branching ratio $BR(H \rightarrow \psi\psi)$ as a function of the BDT background efficiency.

It is found that the optimal BDT threshold has some (sometimes large) dependence on the signal lifetime. To balance between performance and complexity, the optimal threshold is determined in spite of the lifetime. The sensitivity is optimised by tuning the BDT threshold according to the different signal mass hypotheses. This is performed initially by finding the BDT threshold that gives the best sensitivity for each signal mass hypothesis. For the vector portal, the signal mass hypotheses under study span a large range of mass from 2 GeV to 20 GeV, and it is observed that there is a larger background in the mass region below 5 GeV as shown in Fig. 8.18. The larger background is contributed by the decays of B mesons, which have a mass of 5 GeV. Hence the BDT threshold that gives

10^{-5} background efficiency (BDT > 0.999) is chosen for $m < 5$ GeV, and the threshold that gives 10^{-4} background efficiency (BDT > 0.997) is chosen for $m > 5$ GeV.

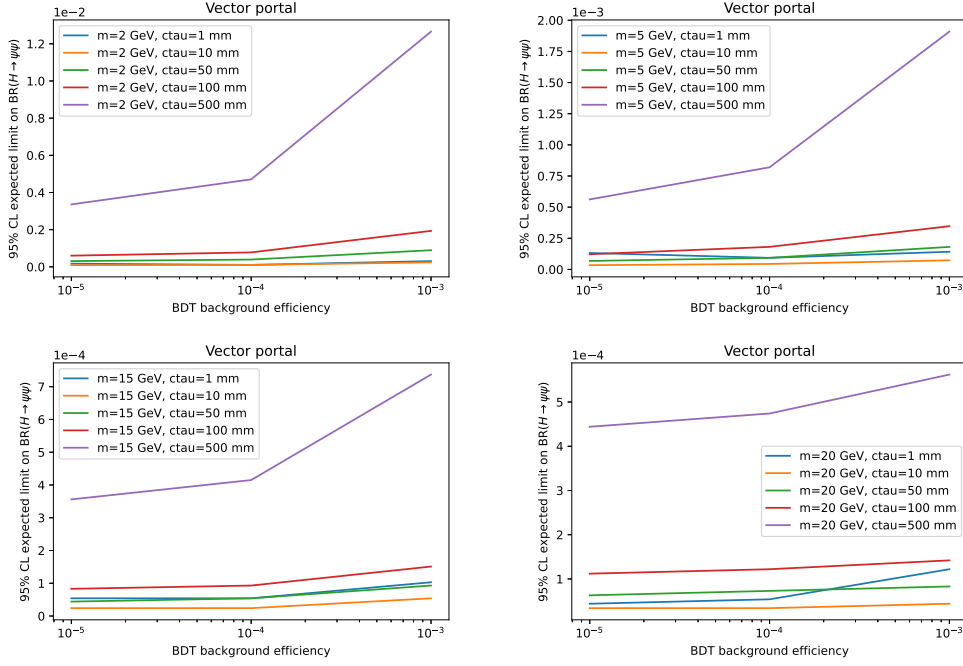


Figure 8.20: 95% confidence level expected upper limits against the BDT background efficiency for $m = 2, 5, 15$ and 20 GeV of the vector portal model. Representative lifetime models are studied for each mass hypothesis.

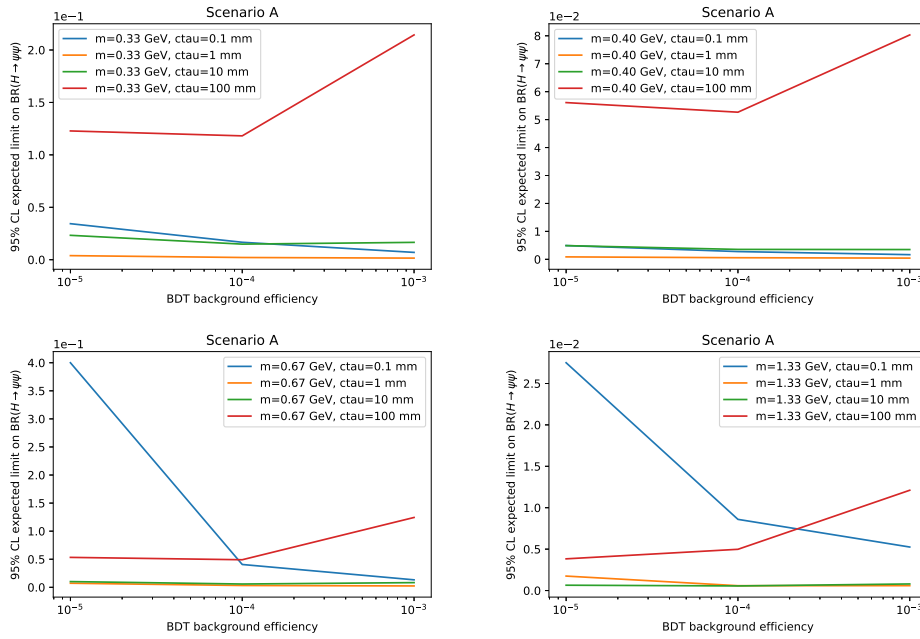


Figure 8.21: 95% confidence level expected upper limits against the BDT background efficiency for $m = 0.33, 0.4, 0.67$ and 1.33 GeV of the scenario A model. Representative lifetime models are studied for each mass hypothesis.

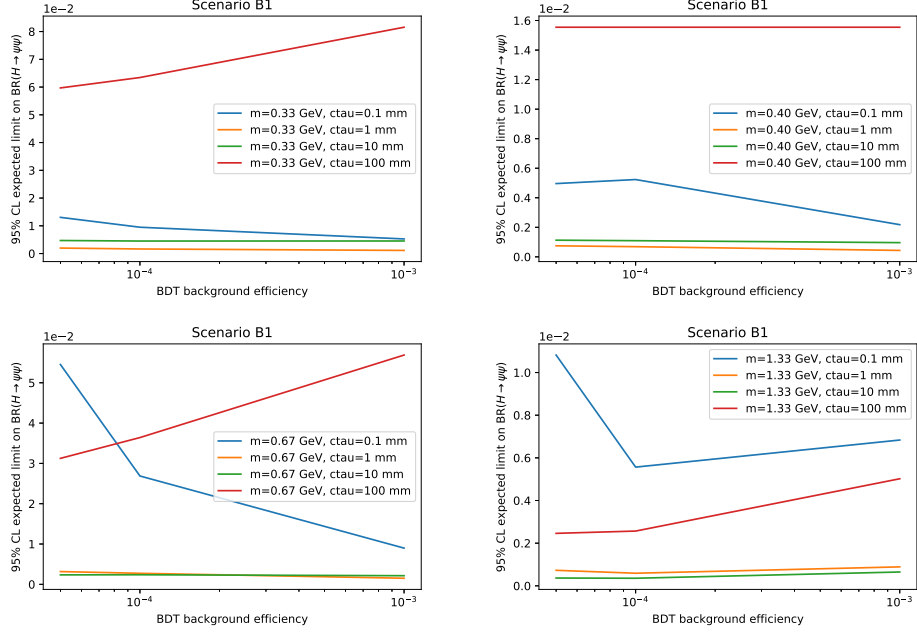


Figure 8.22: 95% confidence level expected upper limits against the BDT background efficiency for $m = 0.33, 0.4, 0.67$ and 1.33 GeV of the scenario B1 model. Representative lifetime models are studied for each mass hypothesis

The BDT thresholds are also optimised for the scenario A and the scenario B1 BDTs. The signal mass hypotheses considered are all below 2 GeV for these models. To balance between performance and complexity, it is found that in general a BDT background efficiency of 10^{-4} is optimal for both scenario A and scenario B1. The corresponding BDT thresholds are $\text{BDT} > 0.98$ for scenario A and $\text{BDT} > 0.991$ for scenario B1. The optimised BDT thresholds are summarised in Table 8.2 for the different mass models in the vector portal. For scenario A and scenario B1, the BDT background efficiency working point of 10^{-4} is used for all the signal masses considered, which are $m_{A'} = 0.33, 0.4, 0.67$ and 1.33 GeV.

<i>Vector portal</i>	$m < 5$ GeV	$m > 5$ GeV
BDT background efficiency working point	10^{-5}	10^{-4}

Table 8.2: The BDT background efficiency working points used for the different mass models in the vector portal after optimisation.

8.6 Systematic uncertainties

8.6.1 Overview

Different sources of systematic uncertainties are considered in the analysis. Most of the uncertainties are concerned with the signal Monte Carlo, as the background estimation

is data-driven. Discrete profiling is used to determine the background uncertainty when fitting with an envelope of functions. For the signal uncertainties, they can come from the modelling of the efficiencies in signal Monte Carlo compared to data, including the uncertainties from the BDT selection, displaced muon identification and the trigger selection. There is also a systematic uncertainty from pileup due to the difference in the pileup distributions between Monte Carlo and data. Other uncertainties originate from measurements from data, including uncertainties on the strong coupling constant, PDF and luminosity. Furthermore, there is a theory uncertainty in the gluon-gluon fusion Higgs production cross-section due to missing higher order terms in the QCD calculation, and variations in the factorisation and renormalisation scales.

8.6.2 Systematic uncertainty from the Boosted Decision Tree

A BDT is used in the analysis to discriminate between the signal and the background by applying a selection based on the BDT output score. The BDT selection efficiency might not be well-modelled compared to what it would be in data. Therefore a scale factor and an associated systematic uncertainty are derived for the BDT selection efficiency in the signal Monte Carlo. The efficiency is studied in the J/ψ mass region in the QCD Monte Carlo and data, which is between 2.91 and 3.27 GeV. The J/ψ resonance is used as a "proxy" for our signal as it is a displaced low-mass resonance that lies within the mass range of the search.

The analysis pre-selections are first applied for events in the J/ψ mass region, which include all the selections described in Section 7.4 except the BDT selection. The BDT score distribution in that mass region for the QCD background Monte Carlo and a subset of data (4.2 fb^{-1}) are shown in Fig 8.23 for the BDT model trained with the vector portal signal model.

A BDT selection is applied to the J/ψ mass region in Monte Carlo and data. The BDT efficiencies in Monte Carlo and data are determined by fitting to the passing and failing events in each case. A double-sided crystal ball function is used to fit the signal and an exponential function is used to fit the background. The fits are shown in Fig. 8.24 and 8.25 for a BDT selection of $\text{BDT} > 0.99$ as an example. The signal efficiencies are obtained by extracting the signal yields in the pass and the fail fits, and then taking the ratio of (pass signal)/(pass signal + fail signal). The fits are performed so that the signal and the background components can be separated, and only the signal component is used in the efficiency calculation.

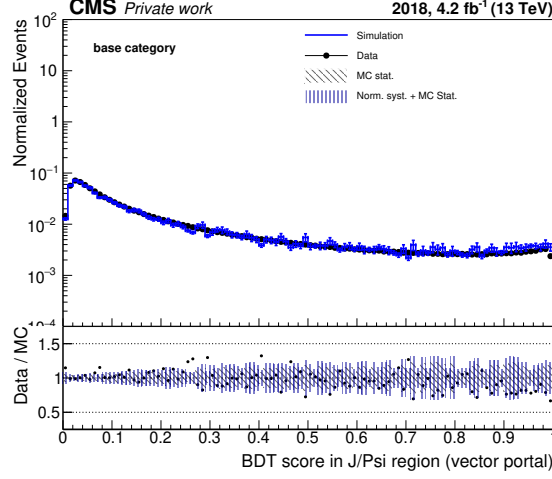


Figure 8.23: BDT score distributions in the J/ψ mass region for QCD Monte Carlo and a subset of data using the BDT trained with the vector portal signal model.

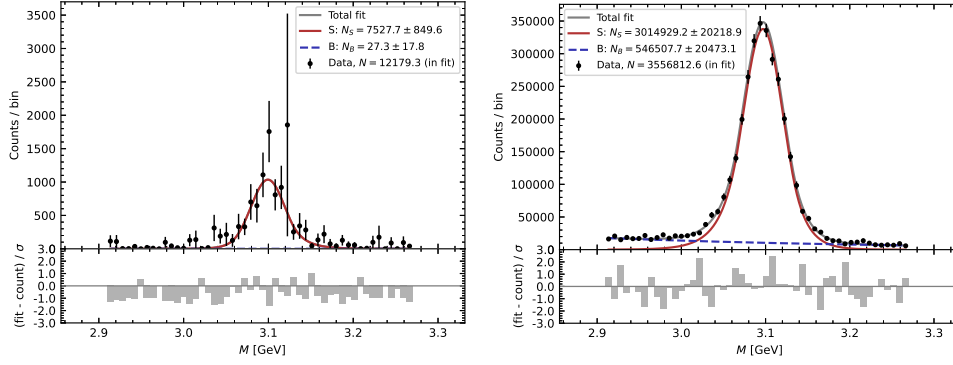


Figure 8.24: Fit for the passing events (left) and the failing events (right) for the QCD Monte Carlo.

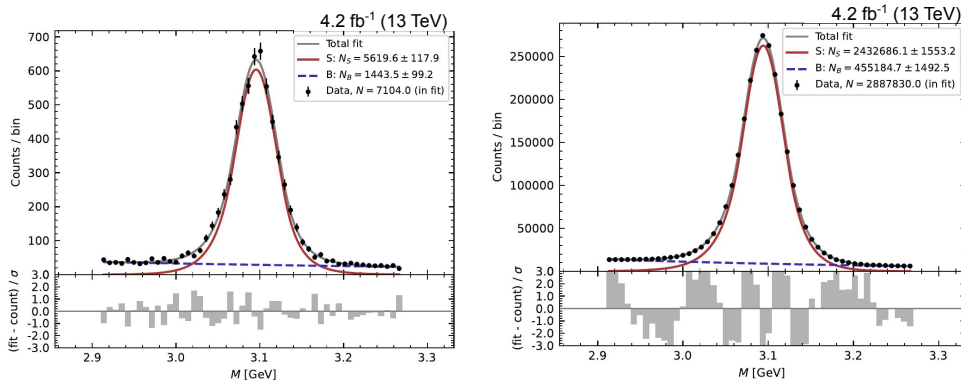


Figure 8.25: Fit for the passing events (left) and the failing events (right) for a subset of data.

The BDT efficiencies are measured for data and Monte Carlo, and the scale factor is derived by taking the ratio between the two efficiencies. The scale factor is found to be

0.93 ± 0.10 for the BDT selection with a background efficiency of 5×10^{-4} .

The BDT efficiencies and scale factors are measured for different BDT thresholds, which correspond to different background efficiencies. The efficiencies and scale factors as a function of the BDT background efficiency is shown in Fig. 8.26 for the vector portal BDT. The BDT scale factors are also measured for the scenario A and scenario B1 BDT, which are shown in Fig. 8.27 and Fig. 8.28 respectively.

No global systematic effect is observed in the BDT scale factors for the different signal models, so a scale factor of unity is assumed, with a systematic uncertainty of 10% to cover any deviations from unity.

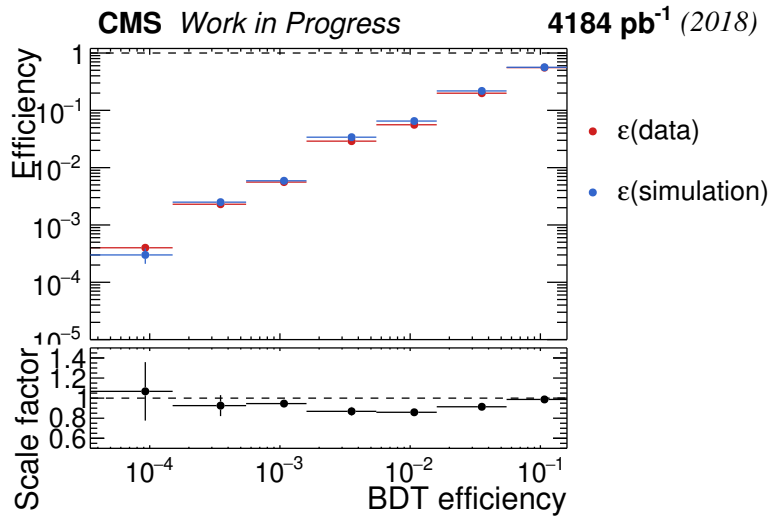


Figure 8.26: BDT efficiencies and scale factors as a function of the background efficiency for the vector portal BDT.

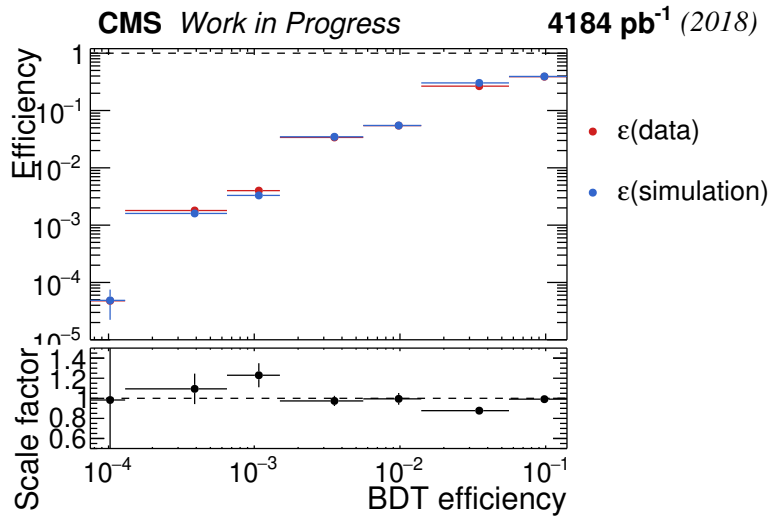


Figure 8.27: BDT efficiencies and scale factors as a function of the background efficiency for the scenario A BDT.

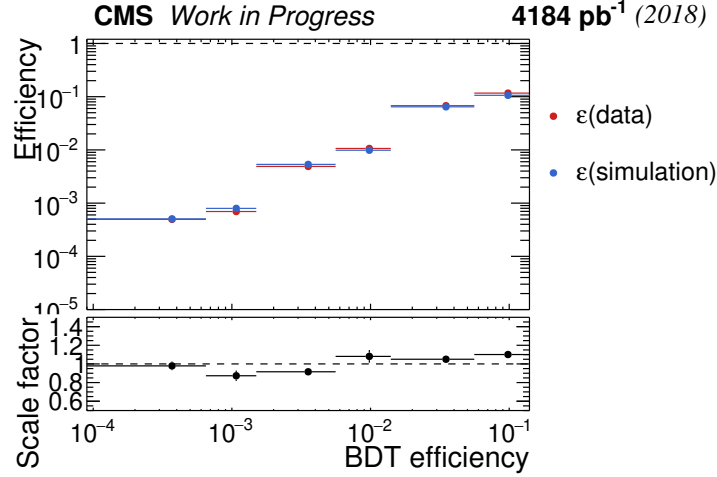


Figure 8.28: BDT efficiencies and scale factors as a function of the background efficiency for the scenario B1 BDT.

8.6.3 Systematic uncertainty from displaced muon identification

In the analysis, muons in the muon secondary vertex are required to pass the loose muon identification. A scale factor and a systematic uncertainty are measured for the muon identification of the displaced muons in order to account for any mis-modelling in the signal Monte Carlo. The "tag-and-probe" technique is implemented for the measurement by first selecting a muon that passes tight quality selections (which is the "tag" muon). It is then paired with a muon of opposite charge (which is the "probe" muon). The invariant mass of the muon pair is required to be between 2.9 and 3.3 GeV, so that it is close to the J/ψ resonance, which is a known resonance in the mass range of the search. The selections used are summarised in Table 8.3. The tag muon is matched to the HLT muon by requiring $\Delta R < 0.1$. From the selected muon pairs, the probes are then effectively unbiased and are used for measurements.

Variables	Cuts
$\text{Prob}(\mu, \mu \text{ vtx})$	> 0.01
Number of hits in the muon chambers (tag μ)	≥ 2
$d_{xy}/\sigma(\text{tag } \mu)$	> 8
$p_T(\text{tag } \mu)$	$> 10 \text{ GeV}$
$p_T(\text{L1 tag } \mu)$	$> 10 \text{ GeV}$
$\Delta R(\text{tag } \mu, \text{L1 tag } \mu)$	< 0.05
Match tag μ with HLT μ	-
$\Delta R(\text{tag } \mu, \text{probe } \mu)$	> 0.15
Muon ID (tag muon)	tight ID
$ \eta(\text{probe muon}) $	< 1.5

Table 8.3: Selections used for the tag-and-probe technique to measure the displaced muon identification efficiency.

A few efficiencies for displaced muons were first considered. Given a muon is reconstructed in the muon chambers, there is the tracker track finding efficiency that takes into account the efficiency of finding a tracker track. In addition, there is the muon track reconstruction efficiency that describes the fraction of tracker tracks that are matched to muons. Finally, the muon identification efficiency describes the fraction of muons that pass the loose identification requirement, given the muon is reconstructed in the tracker. As the tracker track finding efficiency and the track reconstruction efficiency are very close to unity, the dominant inefficiency comes from the muon identification. Therefore only the muon identification efficiency and scale factor are measured, in which the efficiency is defined as:

$$\epsilon_{\text{muon identification}} = \frac{N_{\text{probe}}(\text{loose, tracker muons})}{N_{\text{probe}}(\text{tracker muons})} \quad (8.6)$$

The muon identification efficiency in data and Monte Carlo are derived by fitting the mass distributions of passing and failing dimuon pairs in bins of p_T of the probe muon and l_{xy} of the muon SV. The l_{xy} is chosen because it is used in the analysis for event categorisation. A double-sided crystal ball function is used to fit the J/ψ peak and an exponential function is used to fit the background. Example fits are shown in Figure 8.29. The estimated J/ψ yields are extracted from the fits.

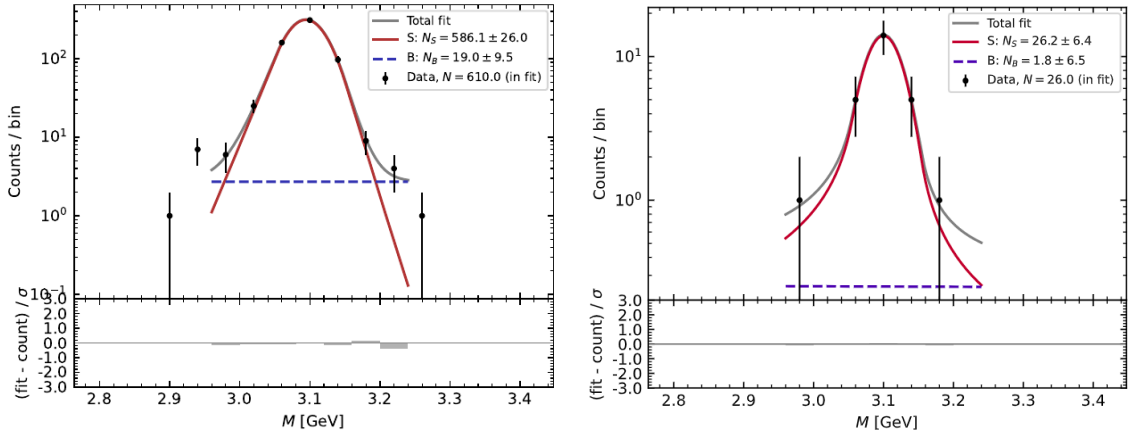


Figure 8.29: Fit for the passing probes (left) and the failing probes (right) for Monte Carlo in the bin of $0.001 \text{ cm} < |d_{xy}| < 0.1 \text{ cm}$ and $4 \text{ GeV} < p_T < 6 \text{ GeV}$.

The muon identification efficiencies are measured for data and Monte Carlo, and scale factors are derived by taking the ratio between the data and Monte Carlo efficiencies. A subset (4.2 fb^{-1}) of the B-parking dataset is used. For Monte Carlo, a sample of $B \rightarrow J/\psi K$ events are used. The efficiencies for data and Monte Carlo, and the scale factors are shown in Fig. 8.30, 8.31 and 8.32 respectively as two-dimensional histograms in bins of p_T and l_{xy} .

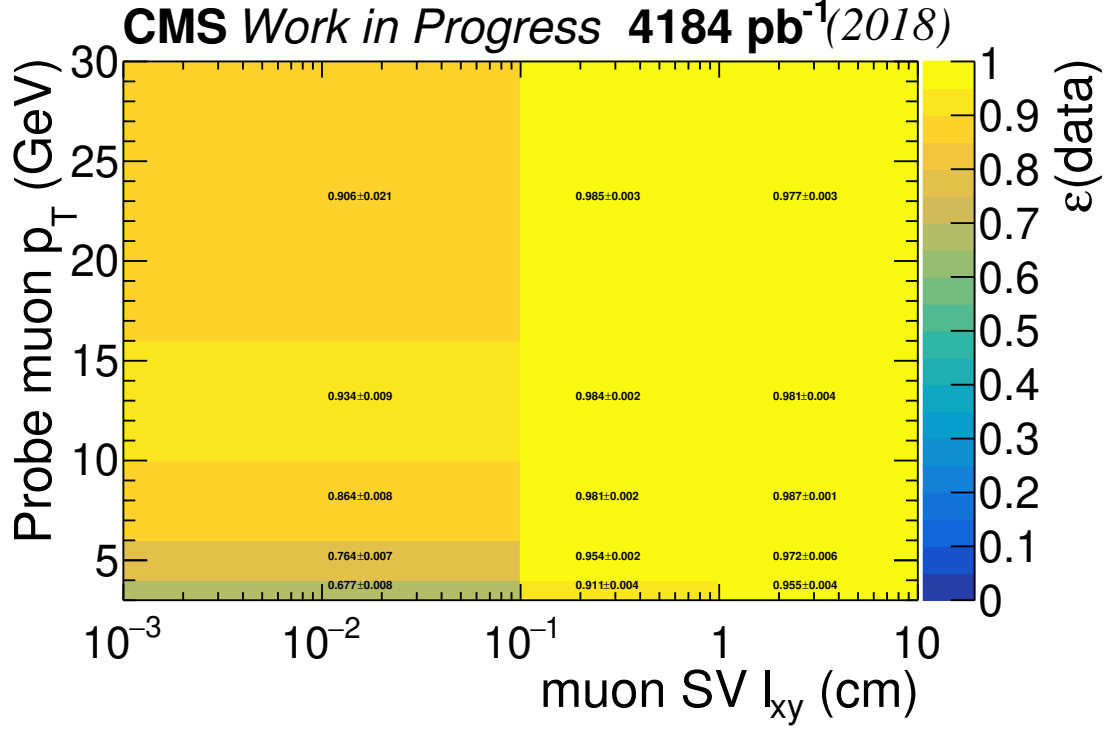


Figure 8.30: Muon identification efficiency for data in bins of transverse momentum and transverse displacement. A subset of B-parking data is used.

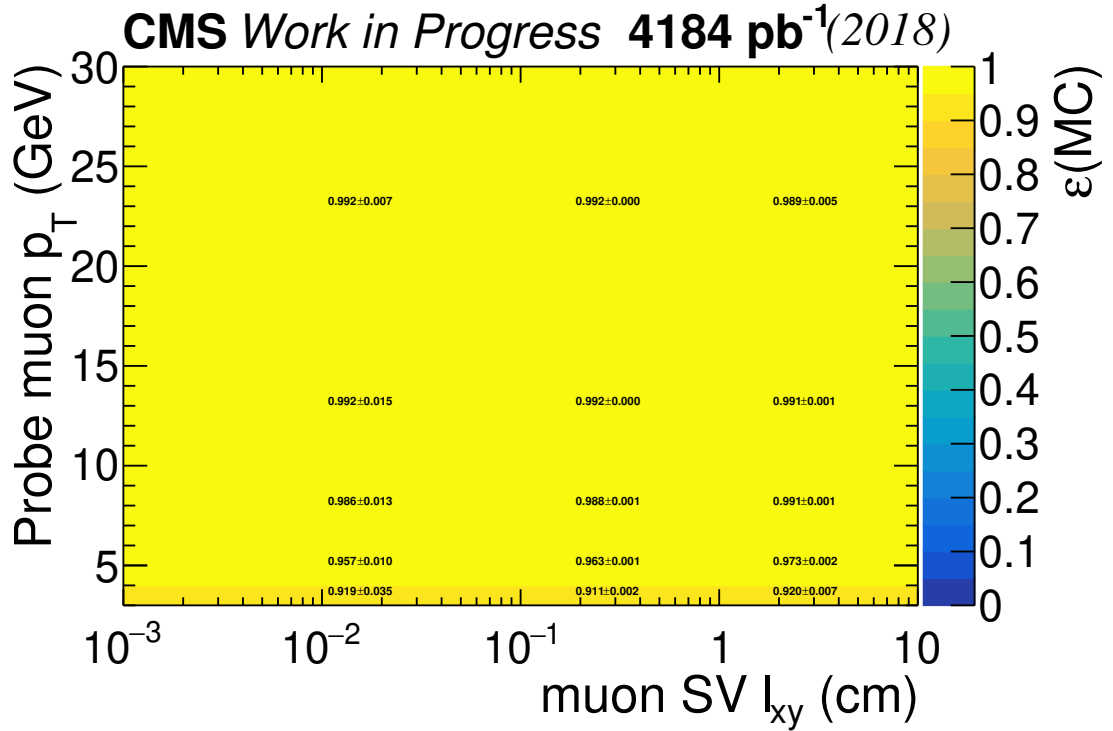


Figure 8.31: Muon identification efficiency for Monte Carlo in bins of transverse momentum and transverse displacement. A Monte Carlo sample of $B \rightarrow J/\psi K$ is used.

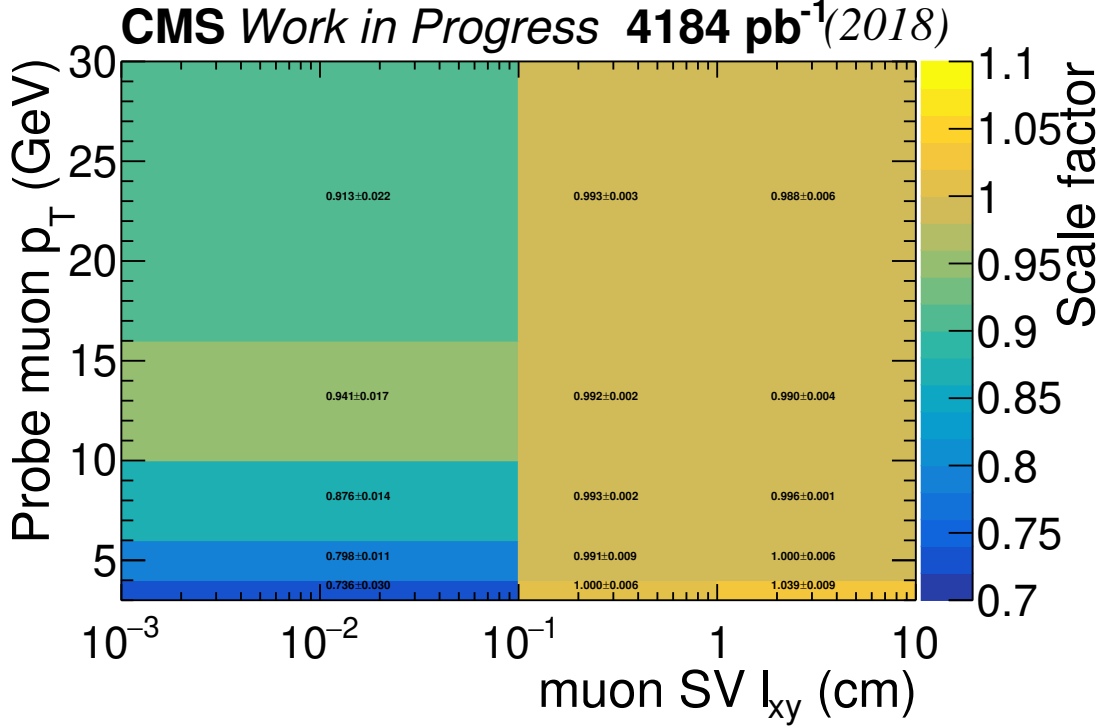


Figure 8.32: Muon identification scale factor in bins of transverse momentum and transverse displacement.

The uncertainties in the results are the combined statistical and systematic uncertainties. The systematic uncertainties are found by varying the mass window of the fit. The fit is first performed with the mass window of [2.96, 3.24] GeV. Then additional fits are performed with the mass window varied up ([2.98, 3.26] GeV) and down ([2.94, 3.22] GeV) respectively to derive a systematic uncertainty on the efficiency obtained by the original fit. The systematic uncertainty is found to be nearly negligible compared to the statistical error.

It is noted that the measurements are statistically limited for $l_{xy} > 10$ cm, as there are very few J/ψ mesons with such high displacement. However the scale factors in the plateau show no strong dependence on p_T nor l_{xy} as seen in Fig. 8.32, and so some assumptions are made. It is assumed that the scale factors are consistent with unity for $l_{xy} > 10$ cm with an uncertainty of 5%, which is a conservative estimate but it is sub-dominant with respect to statistical and other systematic uncertainties.

8.6.4 Systematic uncertainty from the triggers

The B-parking triggers are unique triggers that impose requirements on the impact parameter significance of the muon. Trigger scale factors are first measured using the "tag-and-probe" technique for the HLT_Mu9_IP6 trigger, which accounts for about 80% of the

total luminosity of the dataset.

The tag-and-probe technique is implemented. The tag muon is required to fire the trigger and pass tight quality selections. To reduce the number of combinatorial pairs, some selections are applied on the dimuon pairs. The complete list of selections are summarised in Table 8.4. From the selected pairs, the probes are then effectively unbiased and are used to measure the efficiency of the trigger. The probe muon is classified as passing or failing based on whether it can be matched to the trigger muon.

Variables	Cuts
$\text{Prob}(\mu, \mu \text{ vtx})$	> 0.01
$d_{xy}/\sigma(\text{tag } \mu)$	> 8
$p_T(\text{tag } \mu)$	$> 10 \text{ GeV}$
$p_T(\text{L1 tag } \mu)$	$> 10 \text{ GeV}$
$\Delta R(\text{tag } \mu, \text{L1 tag } \mu)$	< 0.05
Match tag μ with HLT μ	-
$\Delta R(\text{tag } \mu, \text{probe } \mu)$	> 0.15
Muon ID (tag muon)	loose ID
Muon ID (probe muon)	loose ID

Table 8.4: Selections used for the tag-and-probe technique to measure trigger efficiency.

The trigger efficiencies in data and Monte Carlo are derived by fitting the mass distributions of passing and failing dimuon pairs first in bins of p_T and $|d_{xy}|$ of the probe muon. A double-sided crystal ball function is used to fit the J/ψ peak and an exponential function is used to fit the background. Example fits are shown in Fig. 8.33. The estimated J/ψ yields are extracted from the fits. The trigger efficiency is then calculated by:

$$\epsilon_{\text{trigger}} = \frac{N_{\text{passing probes}}}{N_{\text{all probes}}} \quad (8.7)$$

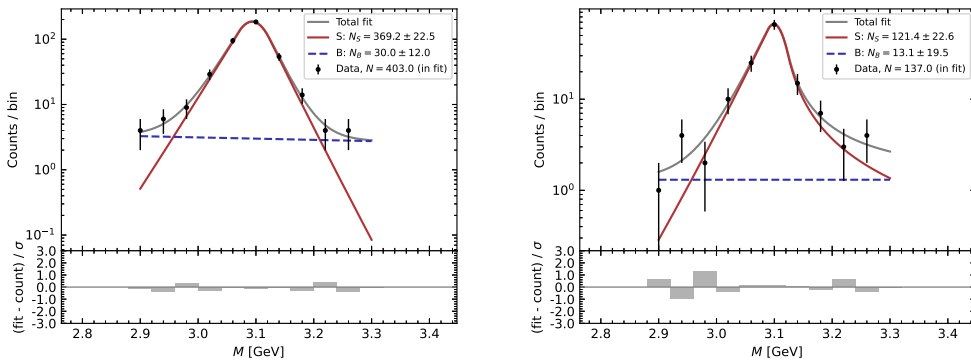


Figure 8.33: Fit for the passing probes (left) and the failing probes (right) for data in the bin of $0.3 \text{ cm} < |d_{xy}| < 0.4 \text{ cm}$ and $10 \text{ GeV} < p_T < 16 \text{ GeV}$.

The trigger efficiencies are measured for data and Monte Carlo, and scale factors are derived by taking the ratio between the data and Monte Carlo. A subset (4.2 fb^{-1}) of the B-parking dataset is used for data, and a sample of $B \rightarrow J/\psi K$ events are used for Monte Carlo. The efficiencies for data and Monte Carlo, and the scale factors are shown in Fig. 8.34, 8.35 and 8.36 respectively as two-dimensional histograms in bins of transverse momentum and transverse impact parameter. A clear efficiency turn-on is observed for both p_T and d_{xy} , especially for p_T where there is a sharp turn-on.

The uncertainties in the results are the combined statistical and systematic uncertainties, where the systematic uncertainties are found by varying the mass window of the fit using the same procedure as described in Section 8.6.3. The systematic uncertainty is found to be nearly negligible compared to the statistical error. The efficiencies and scale factors as a function of the transverse impact parameter of the muon are shown in Fig. 8.37.

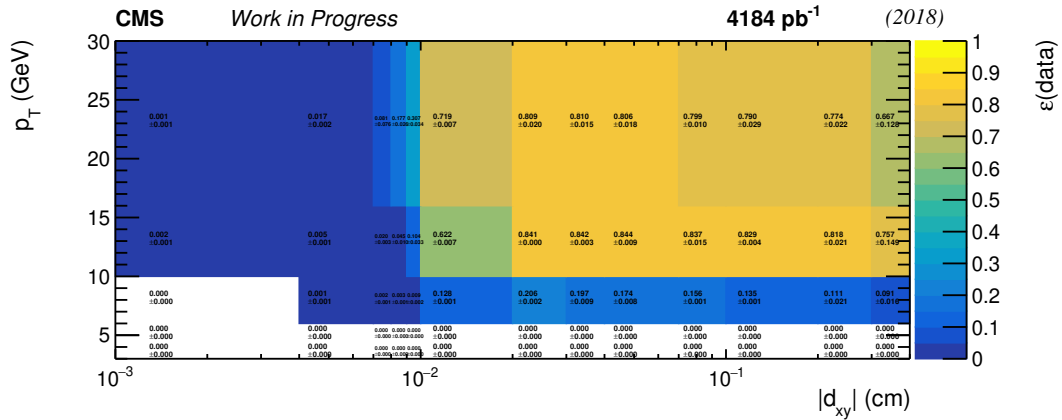


Figure 8.34: Trigger efficiency of the HLT_Mu9_IP6 trigger for data in bins of transverse momentum and transverse impact parameter.

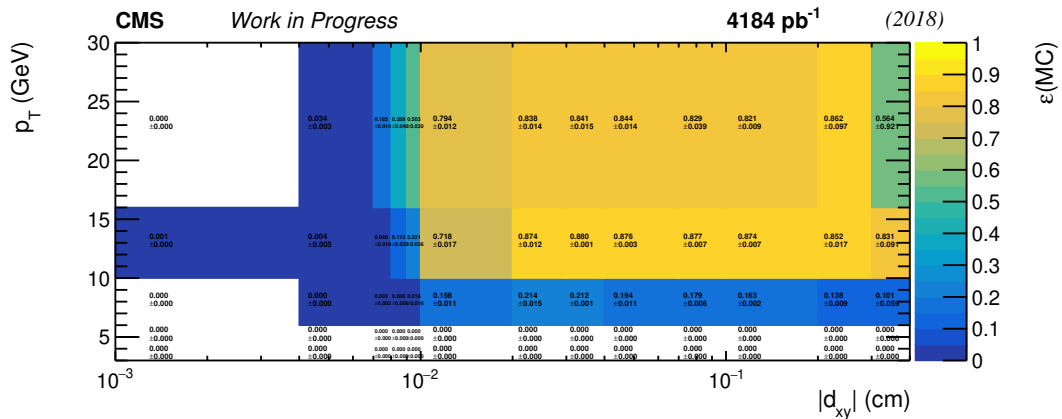


Figure 8.35: Trigger efficiency of the HLT_Mu9_IP6 trigger for Monte Carlo in bins of transverse momentum and transverse impact parameter.

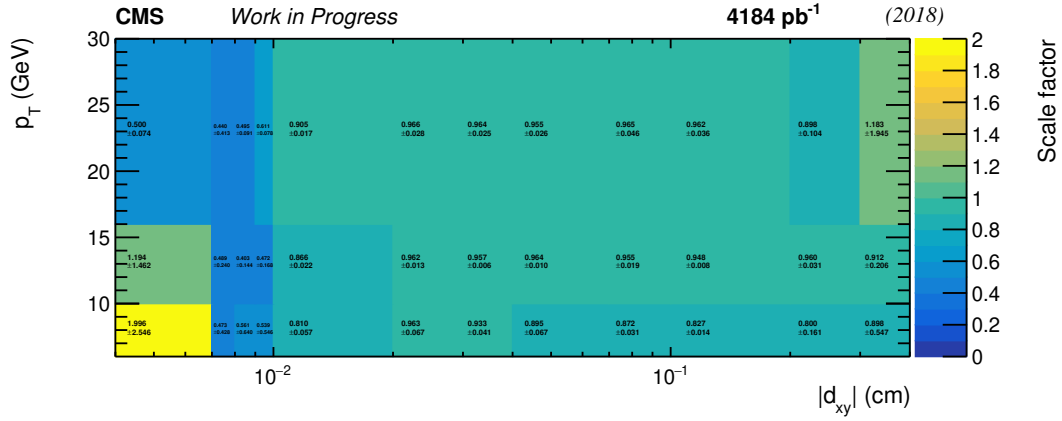


Figure 8.36: Trigger scale factors for the HLT_Mu9_IP6 trigger in bins of transverse momentum and transverse impact parameter.

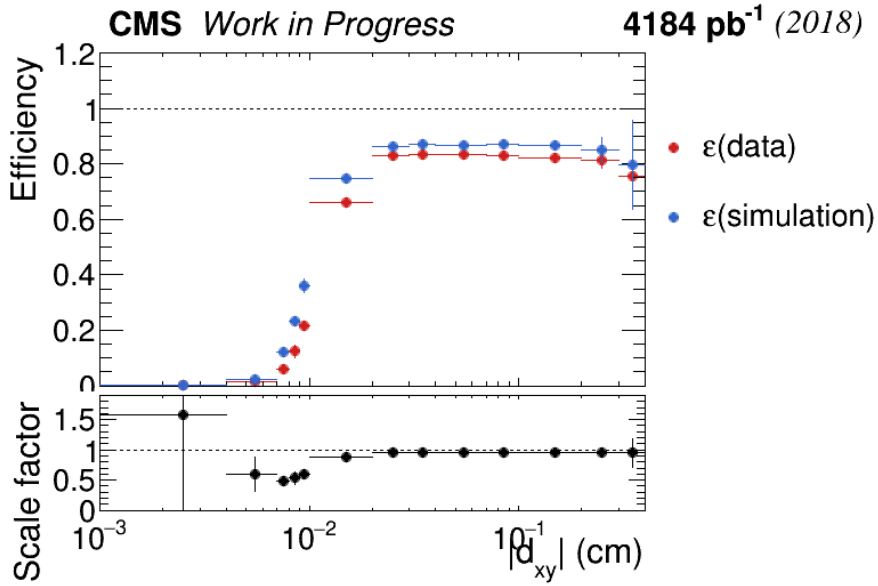


Figure 8.37: Trigger efficiencies and scale factors for the HLT_Mu9_IP6 trigger in bins of transverse impact parameter.

As events are categorised by the transverse displacement of the muon secondary vertex in the analysis, the trigger efficiencies and scale factors are also measured as a function of the transverse displacement. The efficiencies for data and Monte Carlo, and the scale factors are shown in Fig. 8.38, 8.39 and 8.40 respectively as functions of the transverse momentum of the probe muon and the transverse displacement of the associated muon secondary vertex. The efficiencies and scale factors as a function of transverse displacement is shown in Fig. 8.41.

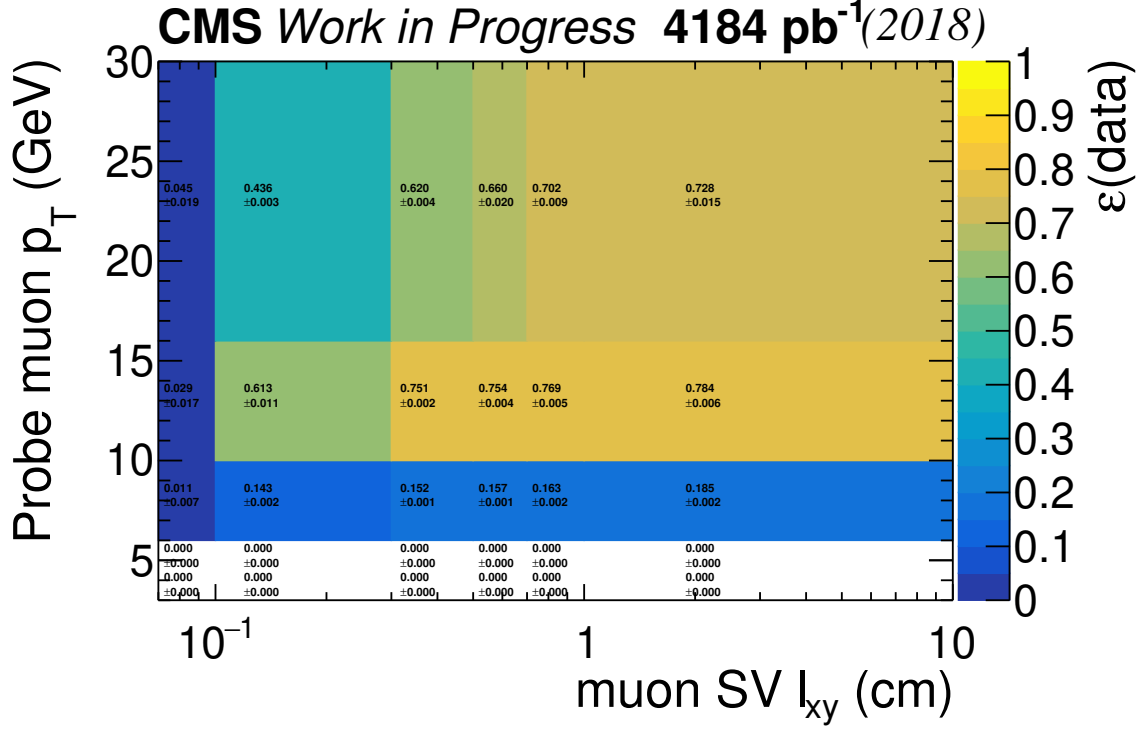


Figure 8.38: Trigger efficiency of the HLT_Mu9_IP6 trigger for data in bins of muon transverse momentum and transverse displacement of the muon secondary vertex.

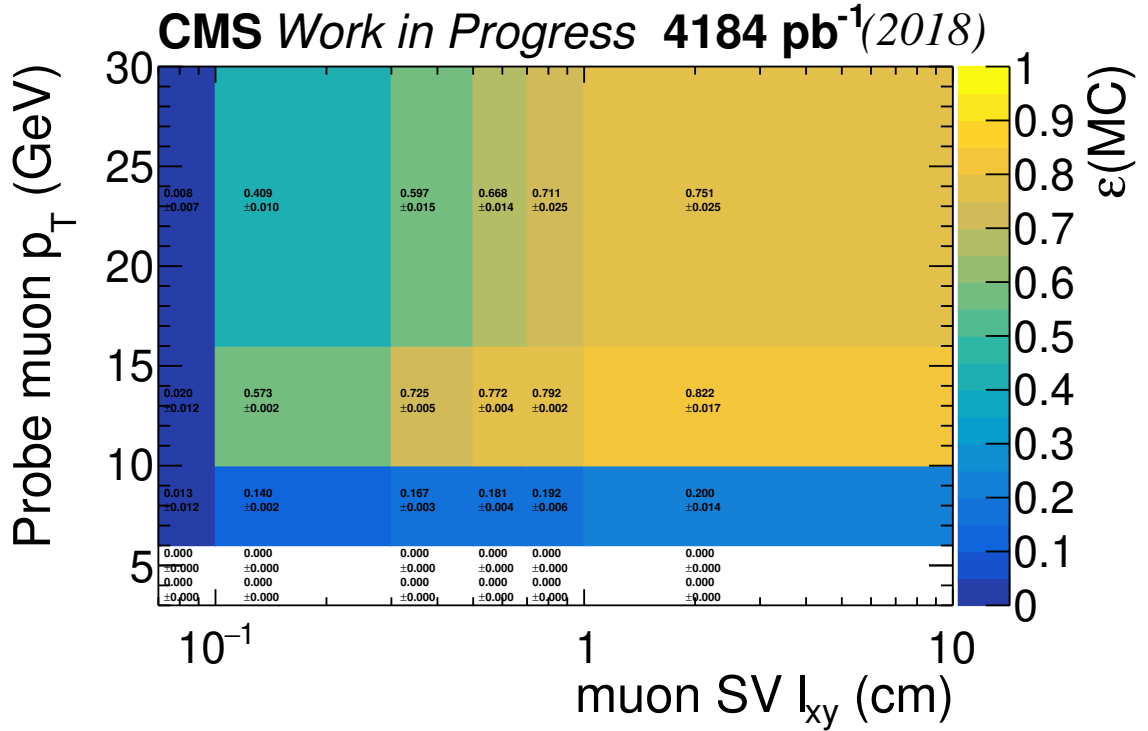


Figure 8.39: Trigger efficiency of the HLT_Mu9_IP6 trigger for Monte Carlo in bins of muon transverse momentum and transverse displacement of the muon secondary vertex.

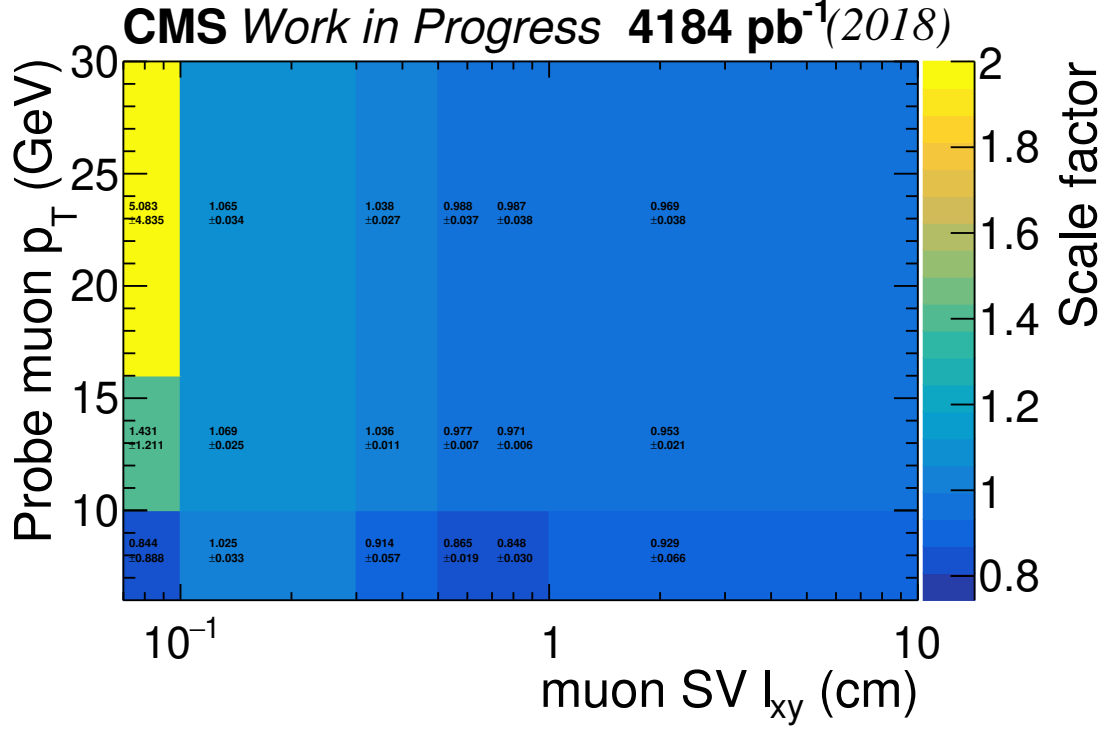


Figure 8.40: Trigger scale factors for the HLT_Mu9_IP6 trigger in bins of muon transverse momentum and transverse displacement of the muon secondary vertex.

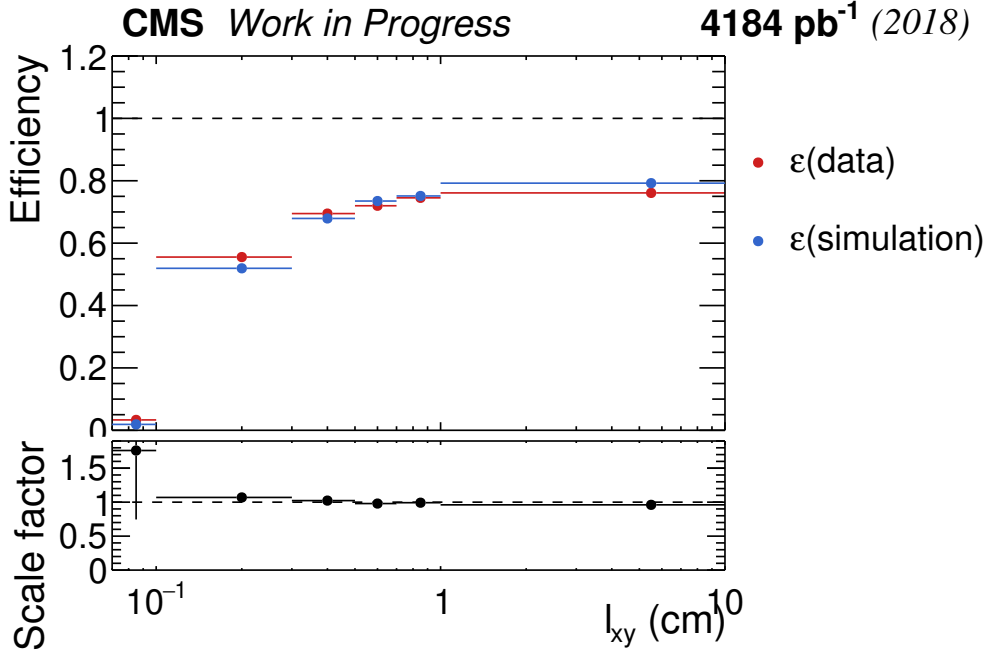


Figure 8.41: Trigger efficiencies and scale factors for the HLT_Mu9_IP6 trigger in bins of transverse displacement of the muon secondary vertex.

For the final analysis, the OR of all B-parking triggers are used, for which the trigger scale factors were measured by the heavy neutrino search [138]. The trigger scale factors

are binned in p_T and the d_{xy} significance ($|d_{xy}/\sigma_{xy}|$) of the probe muon, which are the quantities that are used in the triggers, as shown in Fig. 8.42. These scale factors are applied in the analysis.

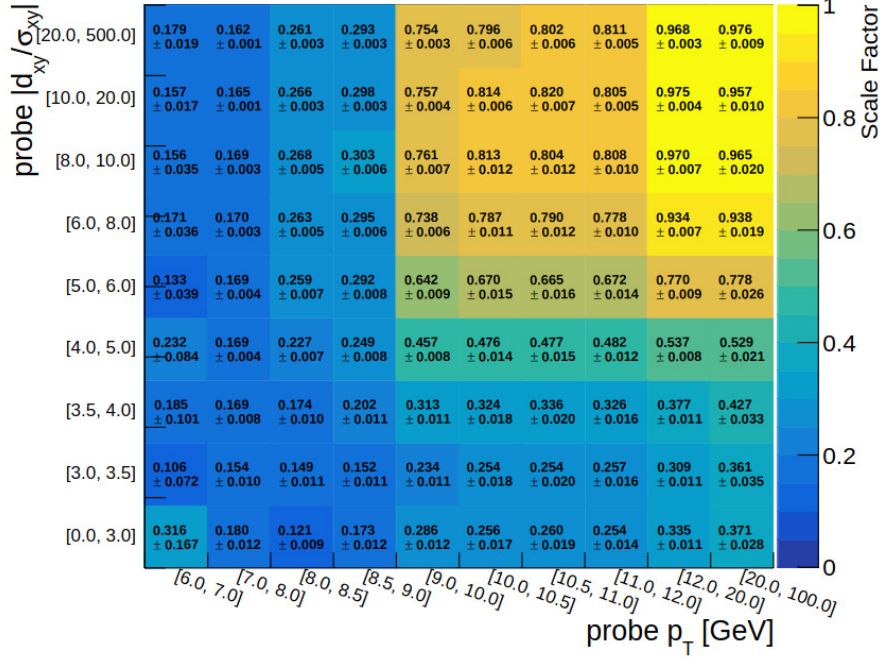


Figure 8.42: Trigger scale factors in bins of transverse momentum and transverse impact parameter significance of the muon for all B-parking triggers [138].

8.6.5 Summary

Various sources of systematic uncertainties are considered in the analysis, which are displaced muon identification, the trigger selection, the BDT selection, pileup, luminosity, the gluon-gluon fusion Higgs production cross-section and discrete profiling. All the systematic uncertainties except the discrete profiling are applied to the signal only, as the background estimation is entirely data driven using the sideband method.

Regarding the BDT selection, the systematic uncertainty on the signal yield is found to be 10%, as discussed in Section 8.6.2.

For the systematic uncertainty on displaced muon identification, scale factors are derived from muon identification efficiencies, as described in Section 8.6.3. The systematic uncertainties on the signal yield at $\pm 1\sigma$ variations of the scale factors are obtained. Only changes that are greater than 1% are included as systematic uncertainties. The resulting systematic uncertainty is typically in the range of 1 – 17% depending on the p_T of the muon and the l_{xy} of the muon secondary vertex.

For the trigger selection, the systematic uncertainties on the signal yield are obtained

using the trigger scale factors presented in Section 8.6.4. Uncertainties in the signal yield that are greater than 1% are included. The resulting systematic uncertainty typically lies between 1 – 2%.

Furthermore, the systematic uncertainty from pileup is obtained by considering the pileup distribution during the time when the B-parking triggers were active. Pileup reweighting is applied to derive the uncertainty at $\pm 1\sigma$. The uncertainty is found to be about 5%.

In addition, the systematic uncertainty coming from the uncertainty in the luminosity measurement is considered. In 2018 the uncertainty was found to be 2.5% [74].

Systematic uncertainties on the gluon-gluon fusion Higgs production cross-section are also considered. These include a theory uncertainty of 3.9%, an uncertainty of 1.9% from the PDF, and an uncertainty of 2.6% from the strong coupling constant [139].

For the background, the systematic uncertainty from discrete profiling when the background is fitted using an envelope of functions is considered. It accounts for the uncertainty coming from the choice of fitting function [134]. The systematic uncertainty is found to be a small compared to the other systematic uncertainties.

Table 8.5 summarises the different sources of systematic uncertainties for the signal. The data statistical uncertainty is found to be about 50%, so the analysis is statistics limited.

	Systematic uncertainty
Displaced muon identification	$\sim 1 - 17\%$
Trigger	$\sim 1 - 2\%$
BDT	10%
Pileup	$\sim 5\%$
Theory uncertainty in ggF Higgs production	3.9%
Strong coupling constant	2.6%
PDF	1.9%
Luminosity	2.5%

Table 8.5: Summary of the systematic uncertainties applied to the signal.

8.7 Results and interpretation

8.7.1 Signal extraction

Maximum likelihood fits are performed under the background-only or background+signal hypotheses, using a sliding mass window for different signal mass hypotheses. The fits are carried out simultaneously for the different categories, using up to twelve categories for each signal hypothesis. Only categories with non-zero signal yield are included. A parametric fit is performed where the functional shapes for the signal and background, as described in Sections 8.3 and 8.4, are used to estimate the signal and background yields in each bin of the dimuon mass distribution. Systematic uncertainties are implemented, as discussed in Section 8.6.5.

The background-only fit is shown in Fig. 8.43 for two mass windows in an example category, together with the signal expected from representative signal models. The p-value of the fit is found to be 0.422 in Fig. 8.43 (top) and 0.719 in Fig. 8.43 (bottom).

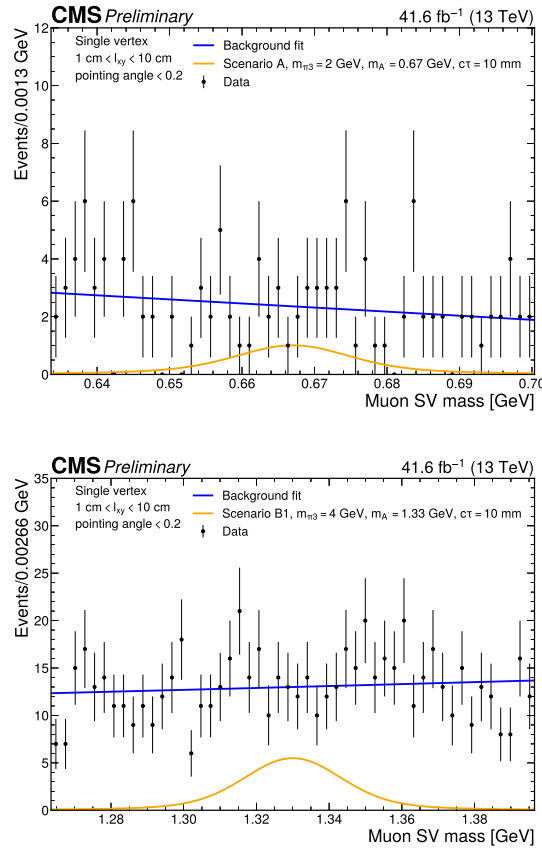


Figure 8.43: Background fit in a mass window centred at 0.67 GeV (top) for scenario A and a mass window centred at 1.33 GeV (bottom) for scenario B1, both in the single vertex category with $1 \text{ cm} < l_{xy} < 10 \text{ cm}$ and pointing angle < 0.2 . The signals expected for a representative scenario A model (top) and a scenario B1 model (bottom) are shown together with the background fits. The signals are normalised to a branching ratio of 0.01.

Some pre-fit and post-fit plots for example categories are shown in Figs. 8.44 and 8.45 for a benchmark model in scenario A and scenario B1 respectively. Goodness-of-fit tests have been performed for the different signal hypotheses in each category, with the results shown in Figs. 8.16 and 8.17.

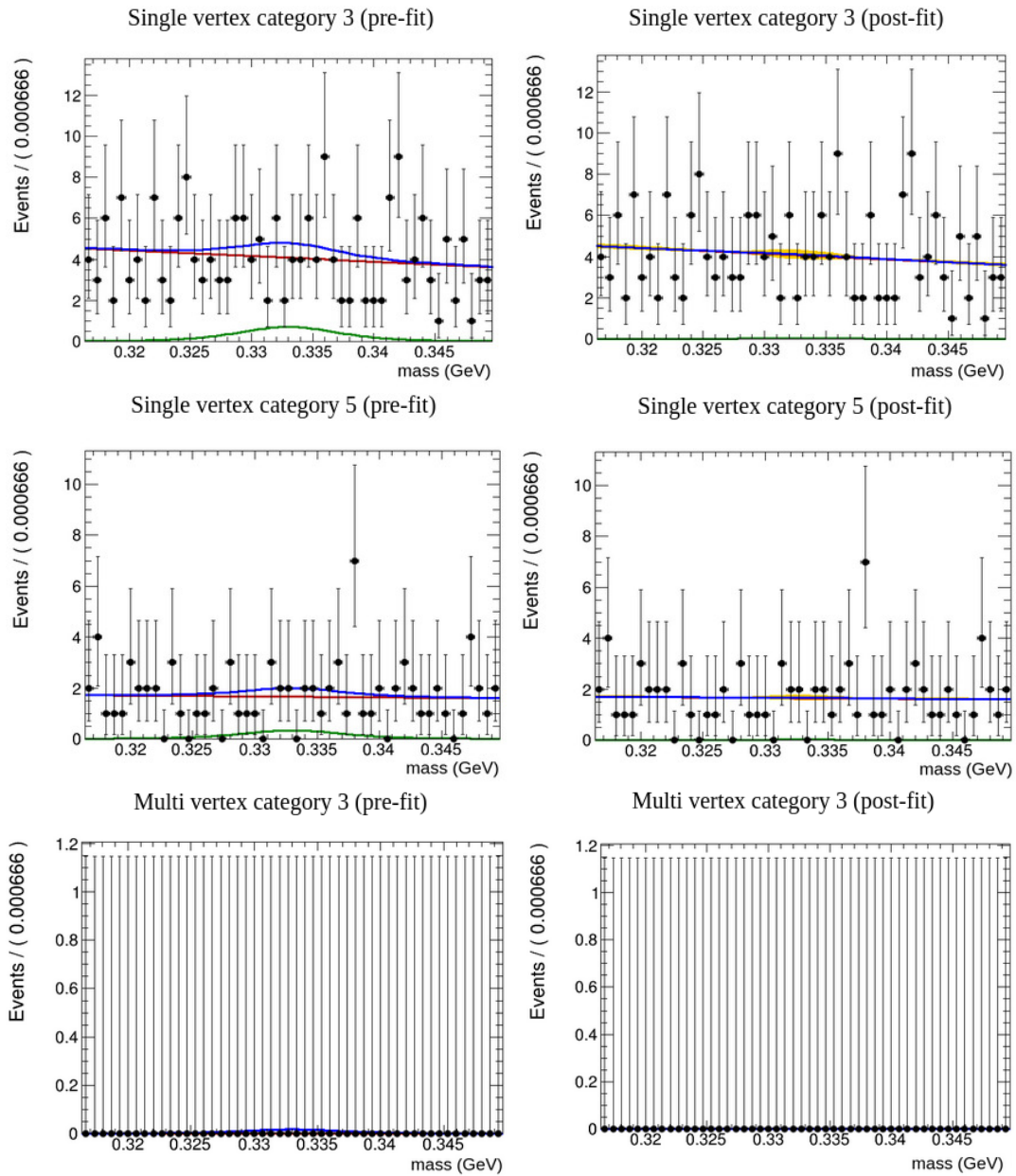


Figure 8.44: Pre-fit and post-fit plots for example categories for scenario A, $m_\pi = 1.0$ GeV, $m_A = 0.33$ GeV, $c\tau = 10$ mm. The signal is shown in green, the background is shown in red, and signal+background is shown in purple. The signal is normalised to a branching ratio of 0.01 in the pre-fit.

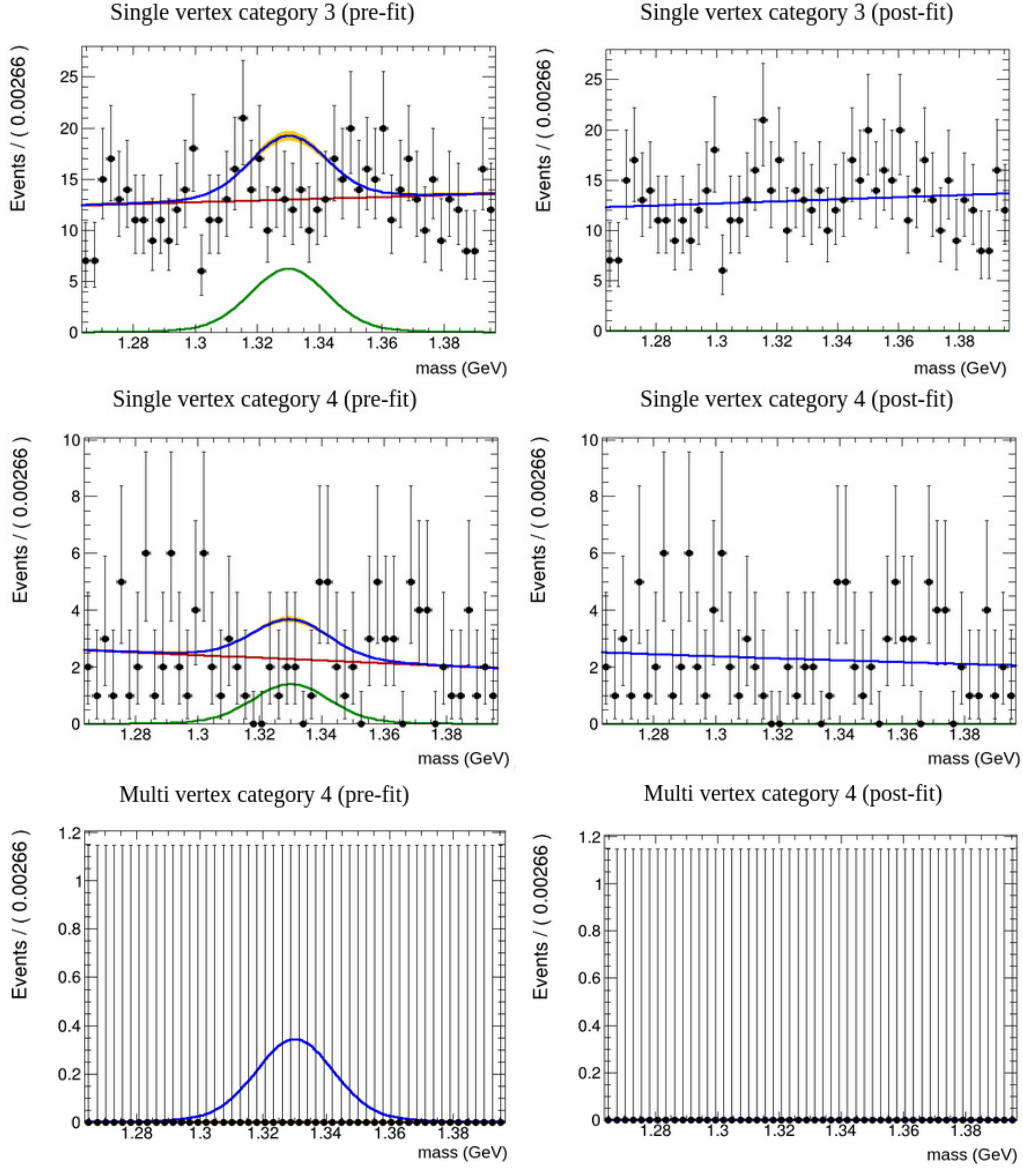


Figure 8.45: Pre-fit and post-fit plots for example categories for scenario B1, $m_{\pi 3} = 4.0$ GeV, $m_A = 1.33$ GeV, $c\tau = 10$ mm. The signal is shown in green, the background is shown in red, and signal+background is shown in purple. The signal is normalised to a branching ratio of 0.01 in the pre-fit.

8.7.2 Branching ratio limits

95% confidence level (CL) upper limits are set on the branching ratio of Higgs decays to dark partons. The limits are produced for different mass and lifetime hypotheses of the dark shower signal models. No significant excess is observed beyond the Standard Model expectation. For the vector portal, the upper limits on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ as a function of the signal lifetime are shown in Fig. 8.46 for a few representative mass hypotheses. The limits on the branching ratio are in the range of 10^{-3} to 10^{-5} over much of phase space.

The corresponding limits for the scenario A model are shown in Fig. 8.47 as a function of lifetime for some representative mass hypotheses. The mass models chosen are in the low-mass regime (below 2 GeV). The limits are in the range of 10^{-2} to 10^{-4} , which show good sensitivity for the low mass models.

The limits for the scenario B1 model are shown in Fig. 8.48 as a function of lifetime. The scenario B1 models are non-pointing, and the model points that are chosen are low-mass. The limits are found to be comparable to those for the scenario A model, which indicates that good sensitivity is achieved for both non-pointing and pointing models.

The constraints are more stringent in general for lower lifetime hypotheses due to the higher secondary vertex reconstruction efficiency at smaller displacement. At very low lifetime, this effect is lessened by the lower signal efficiency of the model. The limits also tend to be more stringent at higher mass due to the lower background present.

It is noted that the observed limits obtained at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ is considerably lower than the observed upper limit set by previous measurements on the invisible branching ratio of the Higgs at 95% CL, $\mathcal{B}(H \rightarrow \text{inv}) < 0.18$ [54].

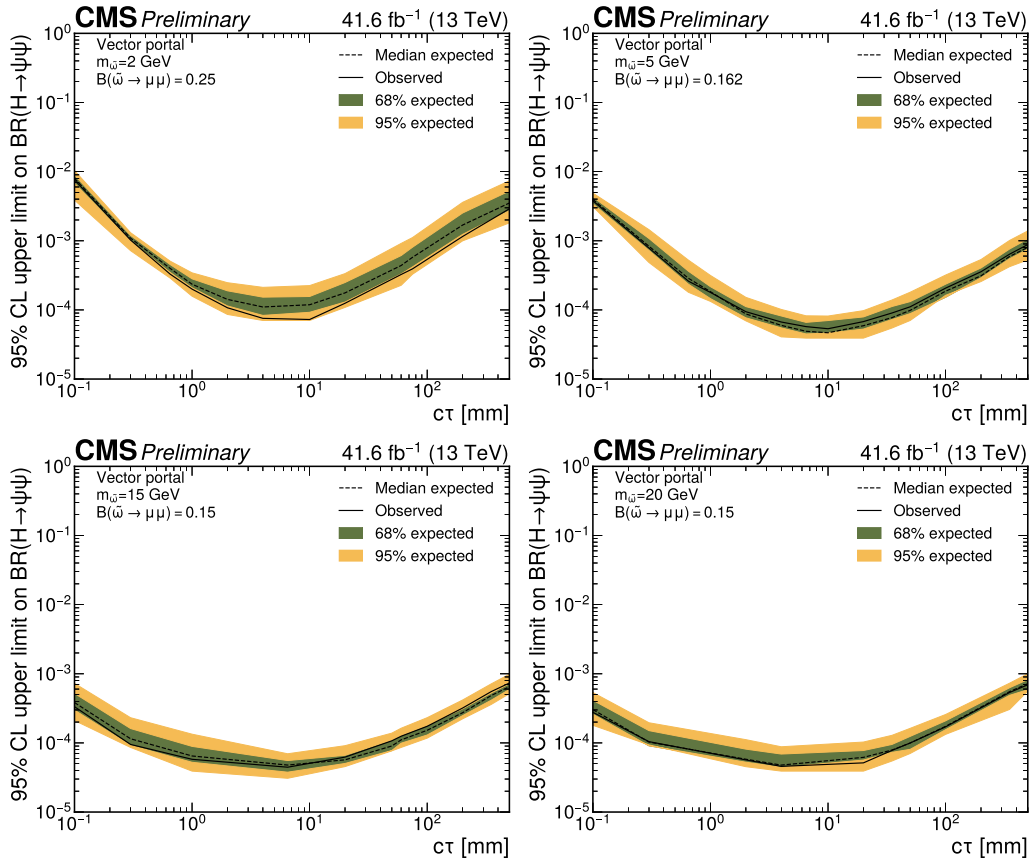


Figure 8.46: Exclusion limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ for representative mass hypotheses for the vector portal model. It is assumed that $m_{\tilde{\omega}} = \tilde{\Lambda} = m_{\tilde{\eta}}$, where $m_{\tilde{\omega}}$, $\tilde{\Lambda}$ and $m_{\tilde{\eta}}$ are the mass of the dark sector spin-one meson, the dark sector confinement scale, and the mass of the dark sector spin-zero meson respectively.

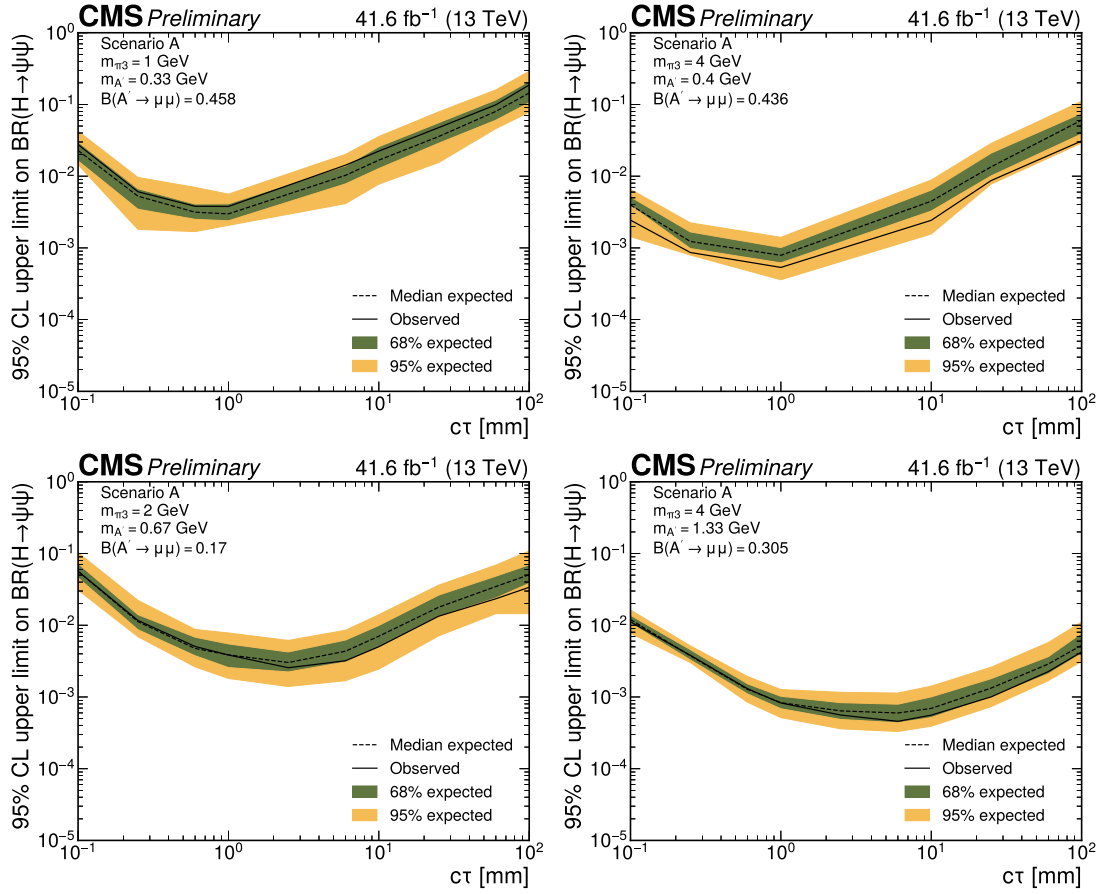


Figure 8.47: Exclusion limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ for representative mass hypotheses for the scenario A model. It is assumed that $m_\eta = \tilde{\Lambda} = 4m_{\pi_2}$ and $\sin \theta = 0.1$, where m_η is the mass of the dark sector pseudoscalar and θ is the mixing angle paramtrising the isospin violation. The branching ratio $\mathcal{B}(\pi_3 \rightarrow A'A')$ is assumed to be one.

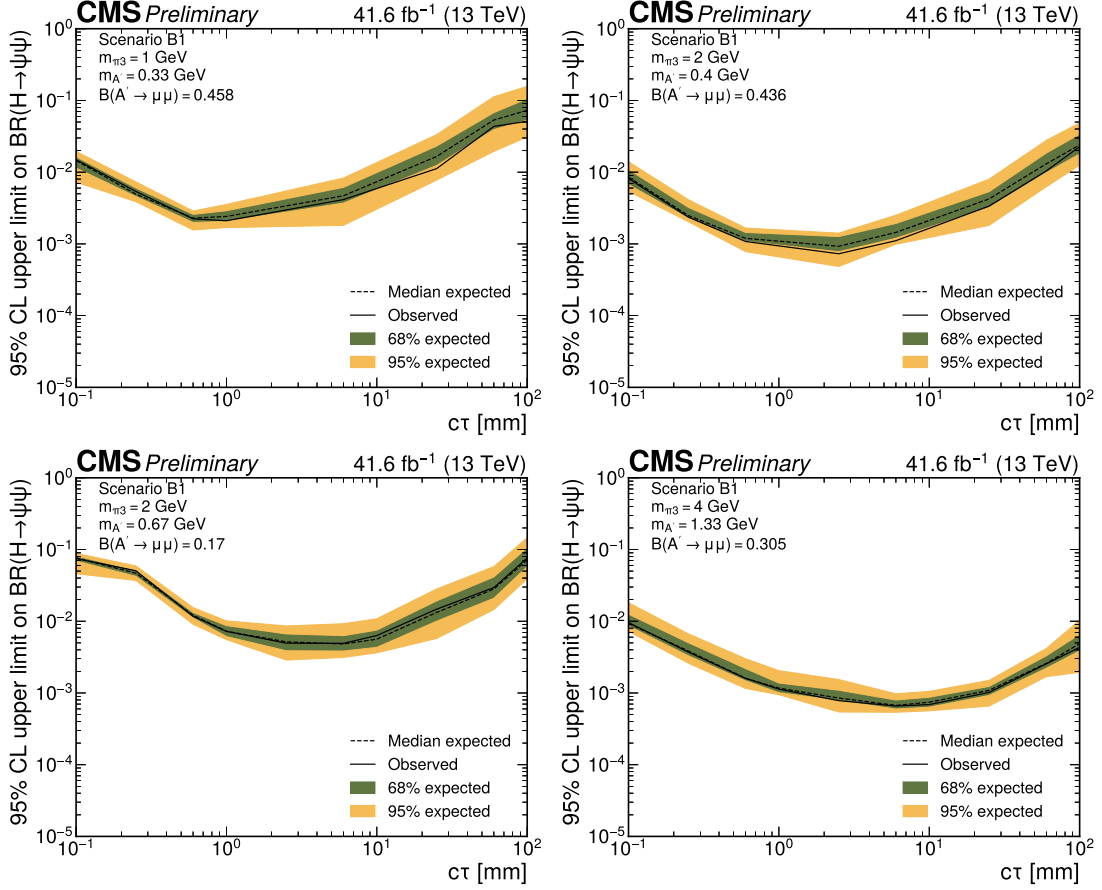


Figure 8.48: Exclusion limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ for representative mass hypotheses for the scenario B1 model. It is assumed that $m_\eta = \tilde{\Lambda} = 4m_{\pi_2}$ and $\sin \theta = 0.1$, where m_η is the mass of the dark sector pseudoscalar and θ is the mixing angle parametrising the isospin violation. The branching ratio $\mathcal{B}(\pi_3 \rightarrow A'A')$ is assumed to be one.

The limits obtained in this search are complementary to those obtained in a previous search performed with the muon detectors of the CMS experiment [140], probing new phase space for lifetime below approximately 0.1 m and mass as low as 2 GeV in the vector portal, where there is an improvement in the limits of up to 2 orders of magnitude. This search makes use of the tracker to look for displaced dimuons, therefore probing lower proper lifetimes. This is complementary to the search in Ref. [140] which looks for hadronic and electromagnetic showers in the muon system, probing the higher lifetime region. As decays into muons very rarely produce a particle shower, the search in Ref. [140] is not sensitive to that decay mode. The sensitivity in this analysis is also improved due to the use of BDT trained for the dark shower model. Apart from the vector portal model, limits are also set on the extended Hidden Valley models scenario A and scenario B1, which are the first constraints imposed on these models.

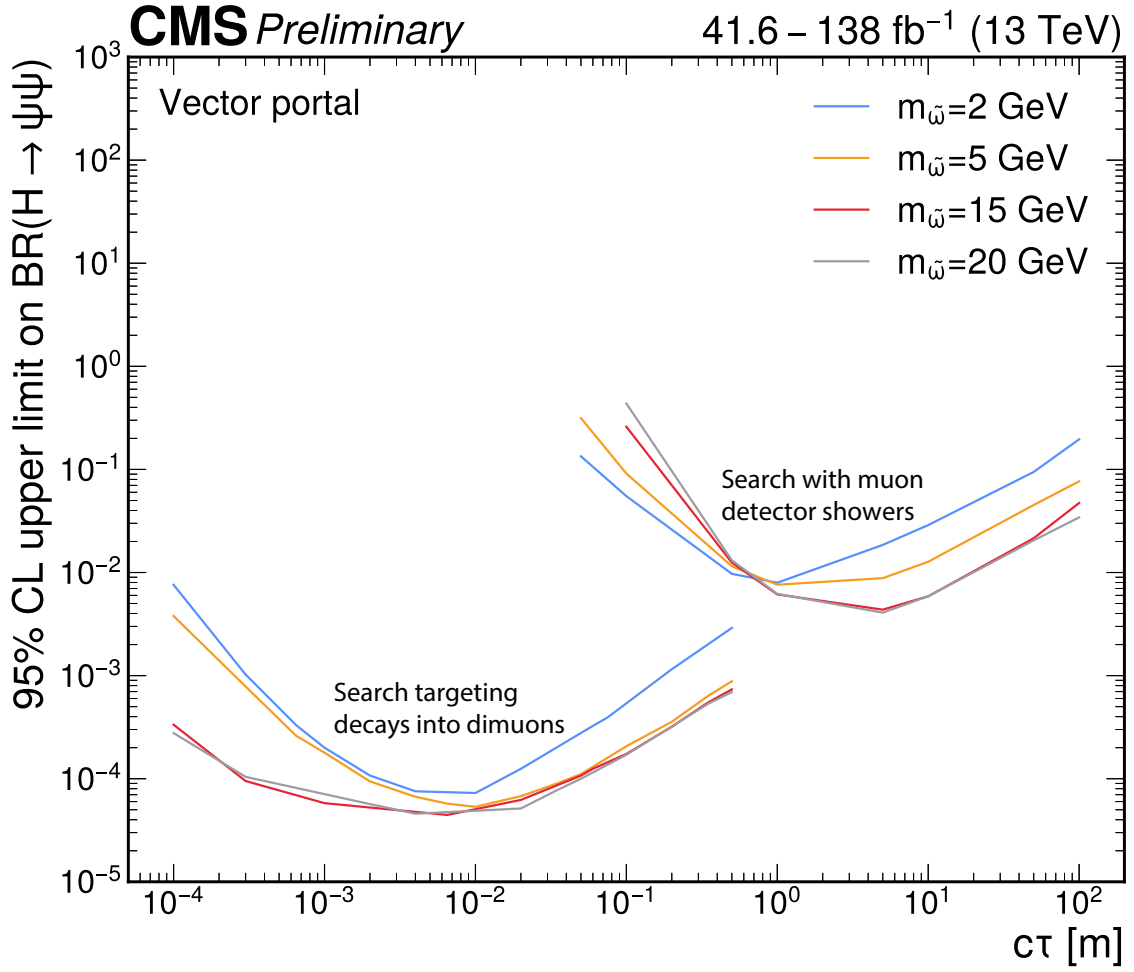


Figure 8.49: Observed limits at 95% CL on the branching ratio $\mathcal{B}(H \rightarrow \psi\psi)$ for the vector portal model from this analysis compared to those from Ref. [140]. This analysis targets decays into dimuons using 41.6 fb⁻¹ of proton-proton collision data, while Ref. [140] is an inclusive search with muon detector showers using 138 fb⁻¹ of proton-proton data that is sensitive to all decay modes except for muons.

Chapter 9

Conclusions

This thesis presents a search for dark showers predicted by the Hidden Valley model through decays into displaced muons at the CMS experiment, which is the first search for this final state at the LHC.

As there is a substantial QCD background, an event-level BDT is trained to distinguish signal from background. A wide variety of event features are used as input variables to the BDT. Muon variables are found to be particularly important for the performance of the BDT, as the signal models are characterised by a large number of displaced muons from the decay of dark showers. The BDT is capable of achieving 10^{-4} background efficiency with higher than 30% signal efficiency.

The analysis considers the pair production of dark partons from decays of the SM Higgs boson, therefore a multi vertex category is used to increase sensitivity. A significant improvement in the signal to background ratio is achieved in the multi vertex category compared to the single vertex category.

A parametric fit is performed on the dimuon invariant mass to extract the signal. Various systematic uncertainties are considered, mainly originating from signal modelling. No significant excess is observed beyond the SM expectation. Upper limits are set on the branching ratio of Higgs decays into dark partons at 95% confidence level. Limits are imposed on the branching ratio down to $\sim 5 \times 10^{-5}$ for the vector portal model, $\sim 6 \times 10^{-4}$ for the scenario A model and $\sim 7 \times 10^{-4}$ for the scenario B1 model. The limits obtained for the vector portal are complementary to a previous search performed with the muon detectors at CMS, probing lifetimes below ~ 1 m and improving the limits by up to 2 orders of magnitude. Similar limits are achieved for scenario A, which is a pointing scenario, and scenario B1, which is a non-pointing scenario. They are the first limits imposed on these extended Hidden Valley models.

When the analysis is published, the likelihood will be provided to allow for recasting. The BDT trained in the analysis could be applied to other signal models that contain displaced muon signatures. Currently the BDT only includes variables of reconstructed objects, but it could be retrained in the future with generator-level variables so that theorists could use it. The impact on the performance of the BDT would need to be investigated.

Bibliography

- [1] 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.* **238** (1980) 471, [doi:10.1086/158003](https://doi.org/10.1086/158003).
- [2] Planck Collaboration, “Planck 2018 results. VI. Cosmological parameters”, *Astron. Astrophys.* **641** (2020) A6, [doi:10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910), [arXiv:1807.06209](https://arxiv.org/abs/1807.06209). [Erratum: *Astron. Astrophys.* 652, C4 (2021)].
- [3] A. Arbey and F. Mahmoudi, “Dark matter and the early Universe: a review”, *Prog. Part. Nucl. Phys.* **119** (2021) 103865, [doi:10.1016/j.ppnp.2021.103865](https://doi.org/10.1016/j.ppnp.2021.103865), [arXiv:2104.11488](https://arxiv.org/abs/2104.11488).
- [4] M. Persic, P. Salucci, and F. Stel, “The Universal rotation curve of spiral galaxies: 1. The Dark matter connection”, *Mon. Not. Roy. Astron. Soc.* **281** (1996) 27, [doi:10.1093/mnras/278.1.27](https://doi.org/10.1093/mnras/278.1.27), [arXiv:astro-ph/9506004](https://arxiv.org/abs/astro-ph/9506004).
- [5] M. Thomson, “Modern particle physics”. Cambridge University Press, New York, 10, 2013. [doi:10.1017/CB09781139525367](https://doi.org/10.1017/CB09781139525367), ISBN 978-1-107-03426-6, 978-1-139-52536-7.
- [6] ATLAS Collaboration, “Combined Measurement of the Higgs Boson Mass from the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ Decay Channels with the ATLAS Detector Using $\sqrt{s}=7, 8$, and 13 TeV pp Collision Data”, *Phys. Rev. Lett.* **131** (2023), no. 25, 251802, [doi:10.1103/PhysRevLett.131.251802](https://doi.org/10.1103/PhysRevLett.131.251802), [arXiv:2308.04775](https://arxiv.org/abs/2308.04775).
- [7] CMS Collaboration, “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30–61, [doi:10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021), [arXiv:1207.7235](https://arxiv.org/abs/1207.7235).
- [8] 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. B*

- 716** (2012) 1–29, [doi:10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020), [arXiv:1207.7214](https://arxiv.org/abs/1207.7214).
- [9] E. Noether, “Invariante variationsprobleme”, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918** (1918) 235–257.
 - [10] V. E. Barnes et al., “Observation of a Hyperon with Strangeness Minus Three”, *Phys. Rev. Lett.* **12** (1964) 204–206, [doi:10.1103/PhysRevLett.12.204](https://doi.org/10.1103/PhysRevLett.12.204).
 - [11] E. D. Bloom et al., “High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees”, *Phys. Rev. Lett.* **23** (1969) 930–934, [doi:10.1103/PhysRevLett.23.930](https://doi.org/10.1103/PhysRevLett.23.930).
 - [12] M. Breidenbach et al., “Observed behavior of highly inelastic electron-proton scattering”, *Phys. Rev. Lett.* **23** (1969) 935–939, [doi:10.1103/PhysRevLett.23.935](https://doi.org/10.1103/PhysRevLett.23.935).
 - [13] K. G. Wilson, “Confinement of Quarks”, *Phys. Rev. D* **10** (1974) 2445–2459, [doi:10.1103/PhysRevD.10.2445](https://doi.org/10.1103/PhysRevD.10.2445).
 - [14] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Nonabelian Gauge Theories”, *Phys. Rev. Lett.* **30** (1973) 1343–1346, [doi:10.1103/PhysRevLett.30.1343](https://doi.org/10.1103/PhysRevLett.30.1343).
 - [15] H. D. Politzer, “Reliable perturbative results for strong interactions?”, *Phys. Rev. Lett.* **30** (Jun, 1973) 1346–1349, [doi:10.1103/PhysRevLett.30.1346](https://doi.org/10.1103/PhysRevLett.30.1346).
 - [16] T. D. Lee and C. N. Yang, “Question of parity conservation in weak interactions”, *Phys. Rev.* **104** (Oct, 1956) 254–258, [doi:10.1103/PhysRev.104.254](https://doi.org/10.1103/PhysRev.104.254).
 - [17] S. L. Glashow, “The renormalizability of vector meson interactions”, *Nucl. Phys.* **10** (1959) 107–117, [doi:10.1016/0029-5582\(59\)90196-8](https://doi.org/10.1016/0029-5582(59)90196-8).
 - [18] A. Salam, “Weak and Electromagnetic Interactions”, *Conf. Proc. C* **680519** (1968) 367–377, [doi:10.1142/9789812795915_0034](https://doi.org/10.1142/9789812795915_0034).
 - [19] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (1967) 1264–1266, [doi:10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
 - [20] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons”, *Phys. Rev. Lett.* **13** (Aug, 1964) 321–323, [doi:10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
 - [21] P. W. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys. Lett.* **12** (1964) 132–133, [doi:10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9).

- [22] P. W. Higgs, “Broken symmetries and the masses of gauge bosons”, *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, [doi:10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [23] P. W. Higgs, “Spontaneous symmetry breakdown without massless bosons”, *Phys. Rev.* **145** (May, 1966) 1156–1163, [doi:10.1103/PhysRev.145.1156](https://doi.org/10.1103/PhysRev.145.1156).
- [24] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global conservation laws and massless particles”, *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, [doi:10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585).
- [25] T. W. B. Kibble, “Symmetry breaking in non-abelian gauge theories”, *Phys. Rev.* **155** (Mar, 1967) 1554–1561, [doi:10.1103/PhysRev.155.1554](https://doi.org/10.1103/PhysRev.155.1554).
- [26] M. Markevitch et al., “Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56”, *Astrophys. J.* **606** (2004) 819–824, [doi:10.1086/383178](https://doi.org/10.1086/383178), [arXiv:astro-ph/0309303](https://arxiv.org/abs/astro-ph/0309303).
- [27] M. J. Strassler and K. M. Zurek, “Echoes of a hidden valley at hadron colliders”, *Phys. Lett. B* **651** (2007) 374, [doi:10.1016/j.physletb.2007.06.055](https://doi.org/10.1016/j.physletb.2007.06.055), [arXiv:hep-ph/0604261](https://arxiv.org/abs/hep-ph/0604261).
- [28] T. Hur, D.-W. Jung, P. Ko, and J. Y. Lee, “Electroweak symmetry breaking and cold dark matter from strongly interacting hidden sector”, *Physics Letters B* **696** (2011), no. 3, 262, [doi:10.1016/j.physletb.2010.12.047](https://doi.org/10.1016/j.physletb.2010.12.047).
- [29] G. D. Kribs, T. S. Roy, J. Terning, and K. M. Zurek, “Quirky composite dark matter”, *Phys. Rev. D* **81** (2010) 095001, [doi:10.1103/PhysRevD.81.095001](https://doi.org/10.1103/PhysRevD.81.095001).
- [30] H. Beauchesne, E. Bertuzzo, and G. Grilli Di Cortona, “Dark matter in Hidden Valley models with stable and unstable light dark mesons”, *JHEP* **04** (2019) 118, [doi:10.1007/JHEP04\(2019\)118](https://doi.org/10.1007/JHEP04(2019)118), [arXiv:1809.10152](https://arxiv.org/abs/1809.10152).
- [31] A. Francis, R. J. Hudspith, R. Lewis, and S. Tulin, “Dark Matter from Strong Dynamics: The Minimal Theory of Dark Baryons”, *JHEP* **12** (2018) 118, [doi:10.1007/JHEP12\(2018\)118](https://doi.org/10.1007/JHEP12(2018)118), [arXiv:1809.09117](https://arxiv.org/abs/1809.09117).
- [32] E. Bernreuther, F. Kahlhoefer, M. Krämer, and P. Tunney, “Strongly interacting dark sectors in the early Universe and at the LHC through a simplified portal”, *JHEP* **01** (2020) 162, [doi:10.1007/JHEP01\(2020\)162](https://doi.org/10.1007/JHEP01(2020)162), [arXiv:1907.04346](https://arxiv.org/abs/1907.04346).
- [33] Z. Chacko, H.-S. Goh, and R. Harnik, “The Twin Higgs: Natural electroweak breaking from mirror symmetry”, *Phys. Rev. Lett.* **96** (2006) 231802, [doi:10.1103/PhysRevLett.96.231802](https://doi.org/10.1103/PhysRevLett.96.231802), [arXiv:hep-ph/0506256](https://arxiv.org/abs/hep-ph/0506256).

- [34] G. Burdman, Z. Chacko, H.-S. Goh, and R. Harnik, “Folded supersymmetry and the LEP paradox”, *JHEP* **02** (2007) 009, [doi:10.1088/1126-6708/2007/02/009](https://doi.org/10.1088/1126-6708/2007/02/009), [arXiv:hep-ph/0609152](https://arxiv.org/abs/hep-ph/0609152).
- [35] N. Craig, A. Katz, M. Strassler, and R. Sundrum, “Naturalness in the Dark at the LHC”, *JHEP* **07** (2015) 105, [doi:10.1007/JHEP07\(2015\)105](https://doi.org/10.1007/JHEP07(2015)105), [arXiv:1501.05310](https://arxiv.org/abs/1501.05310).
- [36] Y. Bai and P. Schwaller, “Scale of dark QCD”, *Phys. Rev. D* **89** (2014), no. 6, 063522, [doi:10.1103/PhysRevD.89.063522](https://doi.org/10.1103/PhysRevD.89.063522), [arXiv:1306.4676](https://arxiv.org/abs/1306.4676).
- [37] Y. Grossman and D. J. Robinson, “Composite Dirac Neutrinos”, *JHEP* **01** (2011) 132, [doi:10.1007/JHEP01\(2011\)132](https://doi.org/10.1007/JHEP01(2011)132), [arXiv:1009.2781](https://arxiv.org/abs/1009.2781).
- [38] Z. Chacko, P. J. Fox, R. Harnik, and Z. Liu, “Neutrino Masses from Low Scale Partial Compositeness”, *JHEP* **03** (2021) 112, [doi:10.1007/JHEP03\(2021\)112](https://doi.org/10.1007/JHEP03(2021)112), [arXiv:2012.01443](https://arxiv.org/abs/2012.01443).
- [39] S. Renner and P. Schwaller, “A flavoured dark sector”, *JHEP* **08** (2018) 052, [doi:10.1007/JHEP08\(2018\)052](https://doi.org/10.1007/JHEP08(2018)052), [arXiv:1803.08080](https://arxiv.org/abs/1803.08080).
- [40] L. Li and Y. Tsai, “Detector-size Upper Bounds on Dark Hadron Lifetime from Cosmology”, *JHEP* **05** (2019) 072, [doi:10.1007/JHEP05\(2019\)072](https://doi.org/10.1007/JHEP05(2019)072), [arXiv:1901.09936](https://arxiv.org/abs/1901.09936).
- [41] S. Knapen, J. Shelton, and D. Xu, “Perturbative benchmark models for a dark shower search program”, *Phys. Rev. D* **103** (2021) 115013, [doi:10.1103/PhysRevD.103.115013](https://doi.org/10.1103/PhysRevD.103.115013), [arXiv:2103.01238](https://arxiv.org/abs/2103.01238).
- [42] W. Bensalem and D. Stolarski, “Flavor and CP violation from a QCD-like hidden sector”, *JHEP* **02** (2022) 011, [doi:10.1007/JHEP02\(2022\)011](https://doi.org/10.1007/JHEP02(2022)011), [arXiv:2111.05515](https://arxiv.org/abs/2111.05515).
- [43] P. Schwaller, D. Stolarski, and A. Weiler, “Emerging Jets”, *JHEP* **05** (2015) 059, [doi:10.1007/JHEP05\(2015\)059](https://doi.org/10.1007/JHEP05(2015)059), [arXiv:1502.05409](https://arxiv.org/abs/1502.05409).
- [44] D. Curtin and C. B. Verhaaren, “Discovering Uncolored Naturalness in Exotic Higgs Decays”, *JHEP* **12** (2015) 072, [doi:10.1007/JHEP12\(2015\)072](https://doi.org/10.1007/JHEP12(2015)072), [arXiv:1506.06141](https://arxiv.org/abs/1506.06141).
- [45] C. Csaki, E. Kuflik, S. Lombardo, and O. Slone, “Searching for displaced Higgs boson decays”, *Phys. Rev. D* **92** (2015), no. 7, 073008, [doi:10.1103/PhysRevD.92.073008](https://doi.org/10.1103/PhysRevD.92.073008), [arXiv:1508.01522](https://arxiv.org/abs/1508.01522).

- [46] A. Pierce, B. Shakya, Y. Tsai, and Y. Zhao, “Searching for confining hidden valleys at LHCb, ATLAS, and CMS”, *Phys. Rev. D* **97** (2018), no. 9, 095033, [doi:10.1103/PhysRevD.97.095033](https://doi.org/10.1103/PhysRevD.97.095033), [arXiv:1708.05389](https://arxiv.org/abs/1708.05389).
- [47] S. Alipour-Fard, N. Craig, M. Jiang, and S. Koren, “Long Live the Higgs Factory: Higgs Decays to Long-Lived Particles at Future Lepton Colliders”, *Chin. Phys. C* **43** (2019), no. 5, 053101, [doi:10.1088/1674-1137/43/5/053101](https://doi.org/10.1088/1674-1137/43/5/053101), [arXiv:1812.05588](https://arxiv.org/abs/1812.05588).
- [48] S. Alipour-Fard et al., “The second Higgs at the lifetime frontier”, *JHEP* **07** (2020) 029, [doi:10.1007/JHEP07\(2020\)029](https://doi.org/10.1007/JHEP07(2020)029), [arXiv:1812.09315](https://arxiv.org/abs/1812.09315).
- [49] E. Bernreuther et al., “On the challenges of searching for GeV-scale long-lived particles at the LHC”, *JHEP* **04** (2021) 210, [doi:10.1007/JHEP04\(2021\)210](https://doi.org/10.1007/JHEP04(2021)210), [arXiv:2011.06604](https://arxiv.org/abs/2011.06604).
- [50] C. Yuan, H. Zhang, Y. Zhao, and G. Chen, “Producing and detecting long-lived particles at different experiments at the LHC”, *JHEP* **02** (2022) 069, [doi:10.1007/JHEP02\(2022\)069](https://doi.org/10.1007/JHEP02(2022)069), [arXiv:2004.08820](https://arxiv.org/abs/2004.08820).
- [51] L. Carloni and T. Sjostrand, “Visible Effects of Invisible Hidden Valley Radiation”, *JHEP* **09** (2010) 105, [doi:10.1007/JHEP09\(2010\)105](https://doi.org/10.1007/JHEP09(2010)105), [arXiv:1006.2911](https://arxiv.org/abs/1006.2911).
- [52] L. Carloni, J. Rathsmann, and T. Sjostrand, “Discerning Secluded Sector gauge structures”, *JHEP* **04** (2011) 091, [doi:10.1007/JHEP04\(2011\)091](https://doi.org/10.1007/JHEP04(2011)091), [arXiv:1102.3795](https://arxiv.org/abs/1102.3795).
- [53] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159–177, [doi:10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024), [arXiv:1410.3012](https://arxiv.org/abs/1410.3012).
- [54] CMS Collaboration, “Search for invisible decays of the Higgs boson produced via vector boson fusion in proton-proton collisions at $\sqrt{s}=13$ TeV”, *Phys. Rev. D* **105** (2022), no. 9, 092007, [doi:10.1103/PhysRevD.105.092007](https://doi.org/10.1103/PhysRevD.105.092007), [arXiv:2201.11585](https://arxiv.org/abs/2201.11585).
- [55] ATLAS Collaboration, “Combination of searches for invisible decays of the Higgs boson using 139 fb^{−1} of proton-proton collision data at $\sqrt{s}=13$ TeV collected with the ATLAS experiment”, *Phys. Lett. B* **842** (2023) 137963, [doi:10.1016/j.physletb.2023.137963](https://doi.org/10.1016/j.physletb.2023.137963), [arXiv:2301.10731](https://arxiv.org/abs/2301.10731).
- [56] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, “Parton Fragmentation and String Dynamics”, *Phys. Rept.* **97** (1983) 31–145, [doi:10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7).

- [57] E. Witten, “Current Algebra Theorems for the U(1) Goldstone Boson”, *Nucl. Phys. B* **156** (1979) 269–283, [doi:10.1016/0550-3213\(79\)90031-2](https://doi.org/10.1016/0550-3213(79)90031-2).
- [58] T. DeGrand and E. T. Neil, “Repurposing lattice QCD results for composite phenomenology”, *Phys. Rev. D* **101** (2020), no. 3, 034504, [doi:10.1103/PhysRevD.101.034504](https://doi.org/10.1103/PhysRevD.101.034504), [arXiv:1910.08561](https://arxiv.org/abs/1910.08561).
- [59] M. Buschmann, J. Kopp, J. Liu, and P. A. N. Machado, “Lepton Jets from Radiating Dark Matter”, *JHEP* **07** (2015) 045, [doi:10.1007/JHEP07\(2015\)045](https://doi.org/10.1007/JHEP07(2015)045), [arXiv:1505.07459](https://arxiv.org/abs/1505.07459).
- [60] LHCb Collaboration, “Search for $A' \rightarrow \mu^+ \mu^-$ Decays”, *Phys. Rev. Lett.* **124** (2020), no. 4, 041801, [doi:10.1103/PhysRevLett.124.041801](https://doi.org/10.1103/PhysRevLett.124.041801), [arXiv:1910.06926](https://arxiv.org/abs/1910.06926).
- [61] CMS Collaboration, “Search for a narrow resonance decaying to a pair of muons in proton-proton collisions at 13 TeV”,.
- [62] ATLAS Collaboration, “Search for high-mass dilepton resonances using 139 fb⁻¹ of pp collision data collected at $\sqrt{s}=13$ TeV with the ATLAS detector”, *Phys. Lett. B* **796** (2019) 68–87, [doi:10.1016/j.physletb.2019.07.016](https://doi.org/10.1016/j.physletb.2019.07.016), [arXiv:1903.06248](https://arxiv.org/abs/1903.06248).
- [63] D. Curtin, R. Essig, S. Gori, and J. Shelton, “Illuminating Dark Photons with High-Energy Colliders”, *JHEP* **02** (2015) 157, [doi:10.1007/JHEP02\(2015\)157](https://doi.org/10.1007/JHEP02(2015)157), [arXiv:1412.0018](https://arxiv.org/abs/1412.0018).
- [64] T. Cohen, M. Lisanti, and H. K. Lou, “Semivisible Jets: Dark Matter Undercover at the LHC”, *Phys. Rev. Lett.* **115** (2015), no. 17, 171804, [doi:10.1103/PhysRevLett.115.171804](https://doi.org/10.1103/PhysRevLett.115.171804), [arXiv:1503.00009](https://arxiv.org/abs/1503.00009).
- [65] T. Cohen, M. Lisanti, H. K. Lou, and S. Mishra-Sharma, “LHC Searches for Dark Sector Showers”, *JHEP* **11** (2017) 196, [doi:10.1007/JHEP11\(2017\)196](https://doi.org/10.1007/JHEP11(2017)196), [arXiv:1707.05326](https://arxiv.org/abs/1707.05326).
- [66] S. Born, R. Karur, S. Knapen, and J. Shelton, “Scouting for dark showers at CMS and LHCb”, *Phys. Rev. D* **108** (2023) 035034, [doi:10.1103/PhysRevD.108.035034](https://doi.org/10.1103/PhysRevD.108.035034), [arXiv:2303.04167](https://arxiv.org/abs/2303.04167).
- [67] M. J. Strassler and K. M. Zurek, “Discovering the Higgs through highly-displaced vertices”, *Phys. Lett. B* **661** (2008) 263–267, [doi:10.1016/j.physletb.2008.02.008](https://doi.org/10.1016/j.physletb.2008.02.008), [arXiv:hep-ph/0605193](https://arxiv.org/abs/hep-ph/0605193).

- [68] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001, [doi:10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [69] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, [doi:10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [70] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003, [doi:10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [71] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002, [doi:10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [72] LHCb Collaboration, “The LHCb Detector at the LHC”, *JINST* **3** (2008) S08005, [doi:10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [73] “Facts and figures about the LHC”. <https://home.cern/resources/faqs/facts-and-figures-about-lhc>.
- [74] CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV”, technical report, CERN, Geneva, 2019.
- [75] “Luminosity forecasts”. <https://lhc-commissioning.web.cern.ch/schedule/hl-lhc-plots.htm>.
- [76] CMS Collaboration, “The CMS magnet project: Technical Design Report”. Technical design report. CMS. CERN, Geneva, 1997. [doi:10.17181/CERN.6ZU0.V4T9](https://doi.org/10.17181/CERN.6ZU0.V4T9).
- [77] CMS Collaboration, “Cutaway diagrams of CMS detector”,.
- [78] I. Neutelings, “CMS coordinate system”. https://tikz.net/axis3d_cms/.
- [79] CMS Collaboration V. Karimäki, et al., “The CMS tracker system project: Technical Design Report”. Technical design report. CMS. CERN, Geneva, 1997.
- [80] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, [doi:10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009), [arXiv:1405.6569](https://arxiv.org/abs/1405.6569).
- [81] CMS Collaboration, “The CMS Phase-1 Pixel Detector Upgrade”, technical report, CERN, Geneva, 2020.
- [82] R. Fruhwirth and A. Strandlie, “Pattern Recognition, Tracking and Vertex Reconstruction in Particle Detectors”. Particle Acceleration and Detection. Springer Nature, Netherlands, 1st ed. 2021. edition, 2021. ISBN 9783030657710.

- [83] CMS Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report”. Technical design report. CMS. CERN, Geneva, 1997.
- [84] CMS Collaboration, “The CMS ECAL Barrel HV system”, *JINST* **8** (2013) C02039, [doi:10.1088/1748-0221/8/02/C02039](https://doi.org/10.1088/1748-0221/8/02/C02039).
- [85] CMS Collaboration, “The CMS hadron calorimeter project: Technical Design Report”. Technical design report. CMS. CERN, Geneva, 1997.
- [86] CMS Collaboration, “Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV”, *JINST* **15** (2020), no. 05, P05002, [doi:10.1088/1748-0221/15/05/P05002](https://doi.org/10.1088/1748-0221/15/05/P05002), [arXiv:1910.00079](https://arxiv.org/abs/1910.00079).
- [87] CMS Collaboration, J. G. Layter, “The CMS muon project: Technical Design Report”. Technical design report. CMS. CERN, Geneva, 1997.
- [88] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018), no. 06, P06015, [doi:10.1088/1748-0221/13/06/P06015](https://doi.org/10.1088/1748-0221/13/06/P06015), [arXiv:1804.04528](https://arxiv.org/abs/1804.04528).
- [89] CMS Collaboration, “CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems”. Technical design report. CMS.
- [90] CMS Collaboration, S. Cittolin, A. Rácz, and P. Sphicas, “CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project”. Technical design report. CMS. CERN, Geneva, 2002.
- [91] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017), no. 01, P01020, [doi:10.1088/1748-0221/12/01/P01020](https://doi.org/10.1088/1748-0221/12/01/P01020), [arXiv:1609.02366](https://arxiv.org/abs/1609.02366).
- [92] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **15** (2020), no. 10, P10017, [doi:10.1088/1748-0221/15/10/P10017](https://doi.org/10.1088/1748-0221/15/10/P10017), [arXiv:2006.10165](https://arxiv.org/abs/2006.10165).
- [93] Donato, Silvio, “Cms trigger performance”, *EPJ Web Conf.* **182** (2018) 02037, [doi:10.1051/epjconf/201818202037](https://doi.org/10.1051/epjconf/201818202037).
- [94] CMS Collaboration, F. Beaudette, “The CMS Particle Flow Algorithm”, in *International Conference on Calorimetry for the High Energy Frontier*, pp. 295–304. 2013. [arXiv:1401.8155](https://arxiv.org/abs/1401.8155).

- [95] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017), no. 10, P10003, [doi:10.1088/1748-0221/12/10/P10003](#), [arXiv:1706.04965](#).
- [96] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, technical report, CERN, Geneva, 2009.
- [97] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth. A* **262** (1987) 444, [doi:10.1016/0168-9002\(87\)90887-4](#).
- [98] CMS Collaboration, “2017 tracking performance plots”,.
- [99] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *IEEE Proc.* **86** (1998), no. 11, 2210–2239, [doi:10.1109/5.726788](#).
- [100] R. Frühwirth, W. Waltenberger, and P. Vanlaer, “Adaptive Vertex Fitting”, technical report, CERN, Geneva, 2007.
- [101] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, *JINST* **16** (2021), no. 05, P05014, [doi:10.1088/1748-0221/16/05/P05014](#), [arXiv:2012.06888](#).
- [102] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04** (2008) 063, [doi:10.1088/1126-6708/2008/04/063](#), [arXiv:0802.1189](#).
- [103] CMS Collaboration, “Pileup mitigation at CMS in 13 TeV data”, *JINST* **15** (2020), no. 09, P09018, [doi:10.1088/1748-0221/15/09/P09018](#), [arXiv:2003.00503](#).
- [104] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017), no. 02, P02014, [doi:10.1088/1748-0221/12/02/P02014](#), [arXiv:1607.03663](#).
- [105] CMS Collaboration, “Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector”, *JINST* **14** (2019), no. 07, P07004, [doi:10.1088/1748-0221/14/07/P07004](#), [arXiv:1903.06078](#).
- [106] I. Zurbano Fernandez et al., “High-Luminosity Large Hadron Collider (HL-LHC): Technical design report”, [doi:10.23731/CYRM-2020-0010](#).
- [107] CMS Collaboration, “Technical Proposal for the Phase-II Upgrade of the CMS Detector”, [doi:10.17181/CERN.VU8I.D59J](#).

- [108] CMS Collaboration, “The Phase-2 Upgrade of the CMS Level-1 Trigger”, technical report, CERN, Geneva, 2020. LHCC-2020-004.
- [109] CMS Collaboration, “The Phase-2 Upgrade of the CMS Tracker”, technical report, CERN, Geneva, 2017. [doi:10.17181/CERN.QZ28.FLHW](https://doi.org/10.17181/CERN.QZ28.FLHW).
- [110] CMS Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter”, technical report, CERN, Geneva, 2017. [doi:10.17181/CERN.IV8M.1JY2](https://doi.org/10.17181/CERN.IV8M.1JY2).
- [111] C. E. Brown, “Fast Machine Learning in the CMS Level-1 Trigger for the High-Luminosity LHC”, 2023. Presented 25 Jul 2023. <http://cds.cern.ch/record/2875830>.
- [112] R. Tibshirani, “Regression Shrinkage and Selection Via the Lasso”, *J. Roy. Statist. Soc. B* **58** (1996), no. 1, 267–288, [doi:10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x).
- [113] A. E. Hoerl and R. W. Kennard, “Ridge Regression: Biased Estimation for Nonorthogonal Problems”, *Technometrics* **12** (1970), no. 1, 55–67, [doi:10.1080/00401706.1970.10488634](https://doi.org/10.1080/00401706.1970.10488634).
- [114] C. N. Coelho et al., “Automatic heterogeneous quantization of deep neural networks for low-latency inference on the edge for particle detectors”, *Nature Mach. Intell.* **3** (2021) 675–686, [doi:10.1038/s42256-021-00356-5](https://doi.org/10.1038/s42256-021-00356-5), [arXiv:2006.10159](https://arxiv.org/abs/2006.10159).
- [115] C. N. Coelho et al., “Automatic heterogeneous quantization of deep neural networks for low-latency inference on the edge for particle detectors”, *Nature Mach. Intell.* **3** (2021) 675–686, [doi:10.1038/s42256-021-00356-5](https://doi.org/10.1038/s42256-021-00356-5), [arXiv:2006.10159](https://arxiv.org/abs/2006.10159).
- [116] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification”, *JHEP* **10** (2014) 059, [doi:10.1007/JHEP10\(2014\)059](https://doi.org/10.1007/JHEP10(2014)059), [arXiv:1407.6013](https://arxiv.org/abs/1407.6013).
- [117] CMS Collaboration, “Enriching the Physics Program of the CMS Experiment via Data Scouting and Data Parking”, [arXiv:2403.16134](https://arxiv.org/abs/2403.16134).
- [118] CMS Collaboration, “Recording and reconstructing 10 billion unbiased b hadron decays in CMS”, *EPJ Web Conf.* **245** (2020) 01025, [doi:10.1051/epjconf/202024501025](https://doi.org/10.1051/epjconf/202024501025).
- [119] G. Cerminara and B. van Besien, “Automated workflows for critical time-dependent calibrations at the CMS experiment”, *J. Phys. Conf. Ser.* **664** (2015), no. 7, 072009, [doi:10.1088/1742-6596/664/7/072009](https://doi.org/10.1088/1742-6596/664/7/072009).

- [120] C. Bierlich et al., “A comprehensive guide to the physics and usage of PYTHIA 8.3”, *SciPost Phys. Codeb.* **2022** (2022) 8,
[doi:10.21468/SciPostPhysCodeb.8](https://doi.org/10.21468/SciPostPhysCodeb.8), [arXiv:2203.11601](https://arxiv.org/abs/2203.11601).
- [121] A. Buckley et al., “LHAPDF6: parton density access in the LHC precision era”, *Eur. Phys. J. C* **75** (2015) 132, [doi:10.1140/epjc/s10052-015-3318-8](https://doi.org/10.1140/epjc/s10052-015-3318-8),
[arXiv:1412.7420](https://arxiv.org/abs/1412.7420).
- [122] “SM Higgs production cross sections at $\sqrt{s} = 13$ TeV (update in CERN Report4 2016)”, 2016. <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV>.
- [123] T. Chen and C. Guestrin, “XGBoost”, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, aug, 2016. [doi:10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- [124] Mikael Mieskolainen, “ICENET: a deep learning library for HEP”.
<https://github.com/mieskolainen/icenet>.
- [125] P. Fernandez Manteca, “Muon Identification and Isolation efficiencies on Run II data with the CMS experiment”, *PoS LHCP2018* (2018) 068,
[doi:10.22323/1.321.0068](https://doi.org/10.22323/1.321.0068).
- [126] “XGBoost Python API reference”.
https://xgboost.readthedocs.io/en/latest/python/python_api.html.
- [127] T. C. collaboration, “Performance of cms muon reconstruction in pp collision events at $s = 7$ tev”, *Journal of Instrumentation* **7** (oct, 2012) P10002,
[doi:10.1088/1748-0221/7/10/P10002](https://doi.org/10.1088/1748-0221/7/10/P10002).
- [128] CMS Collaboration, “Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV”, technical report, CERN, Geneva, 2010. CMS-PAS-PFT-10-003.
- [129] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435–443,
[doi:10.1016/S0168-9002\(99\)00498-2](https://doi.org/10.1016/S0168-9002(99)00498-2), [arXiv:hep-ex/9902006](https://arxiv.org/abs/hep-ex/9902006).
- [130] A. L. Read, “Presentation of search results: The CL_s technique”, *J. Phys. G* **28** (2002) 2693–2704, [doi:10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [131] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1554,

- [doi:10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0), [arXiv:1007.1727](https://arxiv.org/abs/1007.1727). [Erratum: Eur.Phys.J.C 73, 2501 (2013)].
- [132] A. Wald, “Tests of statistical hypotheses concerning several parameters when the number of observations is large”, *Trans. Amer. Math. Soc.* **54** (1943) 426–482, [doi:10.2307/1990256](https://doi.org/10.2307/1990256).
 - [133] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, *Annals Math. Statist.* **9** (1938), no. 1, 60–62, [doi:10.1214/aoms/1177732360](https://doi.org/10.1214/aoms/1177732360).
 - [134] P. Dauncey, M. Kenzie, N. Wardle, and G. Davies, “Handling uncertainties in background shapes: the discrete profiling method”, *Journal of Instrumentation* **10** (apr, 2015) P04015, [doi:10.1088/1748-0221/10/04/P04015](https://doi.org/10.1088/1748-0221/10/04/P04015).
 - [135] CMS Collaboration, “Search for long-lived particles decaying into muon pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV collected with a dedicated high-rate data stream”, *JHEP* **04** (2022) 062, [doi:10.1007/JHEP04\(2022\)062](https://doi.org/10.1007/JHEP04(2022)062), [arXiv:2112.13769](https://arxiv.org/abs/2112.13769).
 - [136] CMS Collaboration, “Search for a high-mass dimuon resonance produced in association with b quark jets at $\sqrt{s} = 13$ TeV”, *JHEP* **10** (2023) 043, [doi:10.1007/JHEP10\(2023\)043](https://doi.org/10.1007/JHEP10(2023)043), [arXiv:2307.08708](https://arxiv.org/abs/2307.08708).
 - [137] R. D. Cousins, “Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms”, 2013. https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.
 - [138] G. Dissertori, “Search for long-lived heavy neutrinos in B meson decays using the B-parking dataset”, *CMS Analysis Note* **AN-22-138** (2022).
 - [139] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. CERN Yellow Reports: Monographs. CERN, Geneva, 2017. [doi:10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002).
 - [140] CMS Collaboration, “Search for long-lived particles decaying in the CMS muon detectors in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **110** (2024) 032007, [doi:10.1103/PhysRevD.110.032007](https://doi.org/10.1103/PhysRevD.110.032007), [arXiv:2402.01898](https://arxiv.org/abs/2402.01898).
 - [141] D. Guest et al., “Jet Flavor Classification in High-Energy Physics with Deep Neural Networks”, *Phys. Rev. D* **94** (2016), no. 11, 112002, [doi:10.1103/PhysRevD.94.112002](https://doi.org/10.1103/PhysRevD.94.112002), [arXiv:1607.08633](https://arxiv.org/abs/1607.08633).

Appendix A

Details of the B-parking triggers

The B-parking dataset was collected using the following HLT paths, which have the format HLT_MuX_IPY:

- HLT_Mu7_IP4
- HLT_Mu8_IP3
- HLT_Mu8_IP5
- HLT_Mu8_IP6
- HLT_Mu8p5_IP3p5
- HLT_Mu9_IP4
- HLT_Mu9_IP5
- HLT_Mu9_IP6
- HLT_Mu10p5_IP3p5
- HLT_Mu12_IP6

The HLT paths select events with at a muon with $p_T^\mu > X$ GeV and impact parameter significance ($IP_{\text{sig}}^\mu > Y$).

The OR of the following L1T seeds is also required:

- SingleMu22
- SingleMuXer1p5

The first L1T path preselects events with one L1T muon with $p_T^\mu > 22$ GeV. The remaining paths each preselects events with one L1T muon with $p_T^\mu > X$ GeV ($X = 6, 7, 8, 9, 10$,

12, 14, 16 or 18) and $|\eta^\mu| < 1.5$.

All L1 and HLT paths are always present in the trigger menu, but are set to prescale zero by default. As the instantaneous luminosity falls, the set of active trigger paths evolves, with additional paths enabled. Standard data quality criteria are applied to the data for the results. The total integrated luminosity of the full 2018 dataset is 41.9 fb^{-1} .

Appendix B

BDT input variables

The variables that are used as input to the BDT training and the maximum number of objects used are listed in this section. The BDT uses variables from muons, secondary vertices and muon secondary vertices, as shown below:

Muons (16 variables): η , ϕ , p_T , p_T uncertainty, d_{xy} , d_{xy} error, d_z , d_z error, IP , IP significance, charge, tight ID, soft MVA, PF relative isolation (total), mini PF relative isolation (total), index of the associated jet.

Secondary vertices (13 variables): p_T , η , ϕ , x -position, y -position, z -position, d_{xy} , d_{xy} significance, 3D decay length, 3D decay length significance, pointing angle, χ^2 , number of degrees of freedom.

Muon secondary vertices (15 variables): χ^2 , pointing angle, 3D decay length, 3D decay length significance, d_{xy} , d_{xy} significance, $p_T(\mu_1)$, $\eta(\mu_1)$, $\phi(\mu_1)$, $p_T(\mu_2)$, $\eta(\mu_2)$, $\phi(\mu_2)$, vertex x -position, vertex y -position, vertex z -position.

Other variables (3 variables): number of muons, number of secondary vertices and number of muon secondary vertices in the event.

Maximum number of muons used: 8; maximum number of secondary vertices used: 8; maximum number of muon secondary vertices used: 8.

The impact of jet variables on the performance of the BDT is also studied, but they are **not** included in the final training. The variables used are as follow:

Jets (18 variables): p_T , η , ϕ , charged electromagnetic energy fraction, charged hadron energy fraction, neutral electromagnetic energy fraction, neutral hadron energy fraction, muon energy fraction, muon subtraction factor, energy excluded from CHS jets, number of muons in the jet, number of electrons in the jet, number of particles in the jet,

DeepCSV [141] b+bb tag discriminator, quark vs gluon likelihood discriminator, Pileup ID discriminant, index of first matching muon, index of second matching muon.

Maximum number of jets used: 8

Appendix C

BDT hyperparameters

Hyperparameter	Values
num_boost_round	150
booster	'gbtree'
tree_method	'hist'
device	'auto'
learning_rate	0.1
gamma	1.67
max_depth	10
min_child_weight	1.0
max_delta_step	1
subsample	1
colsample_bytree	0.86
colsample_bylevel	0.6
colsample_bynode	0.8
reg_lambda	2.0
reg_alpha	0.05
objective	'binary:logistic'
eval_metric	['logloss']

Table C.1: Hyparameters used for the BDT training

Appendix D

Trigger efficiency against l_{xy} for benchmark signal models

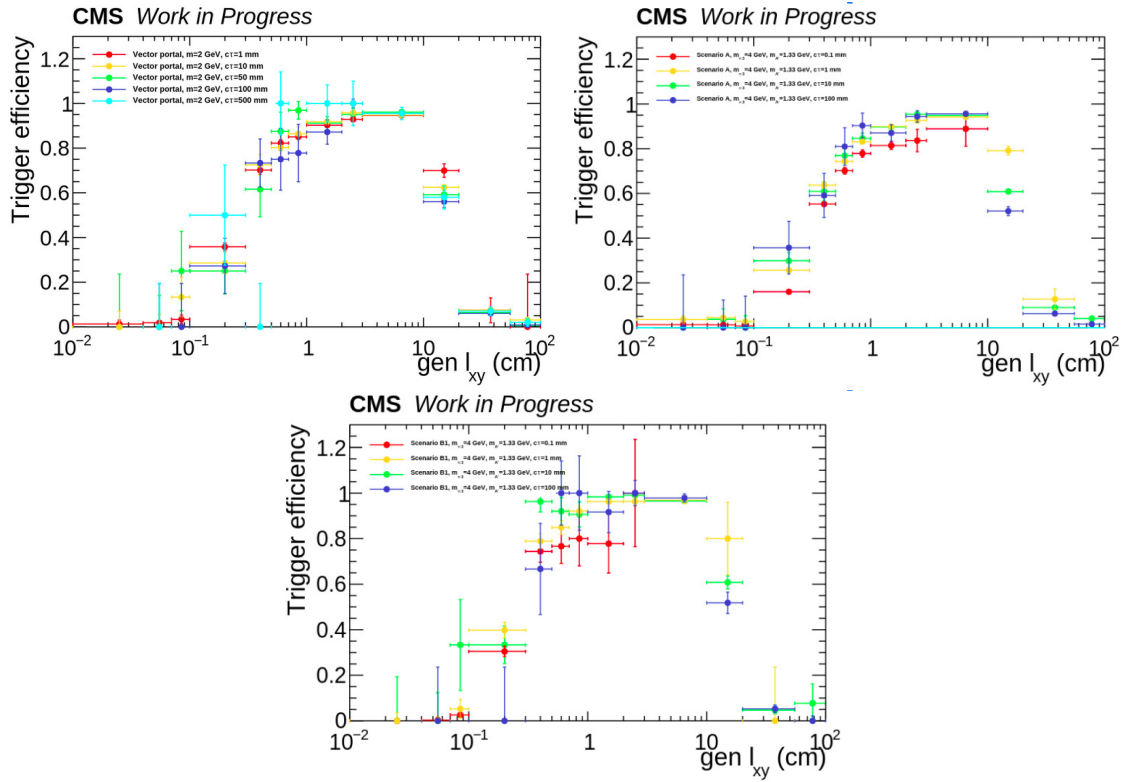


Figure D.1: Trigger efficiency for the HLT_Mu9_IP6 trigger against the transverse displacement l_{xy} of the generated vertex in different signal models. Both muons in the vertex are required to have $p_T > 9$ GeV and $|\eta| < 1.5$. The efficiency drops quickly after about 15 cm.

Appendix E

Selection efficiency against $c\tau$ for benchmark signal models

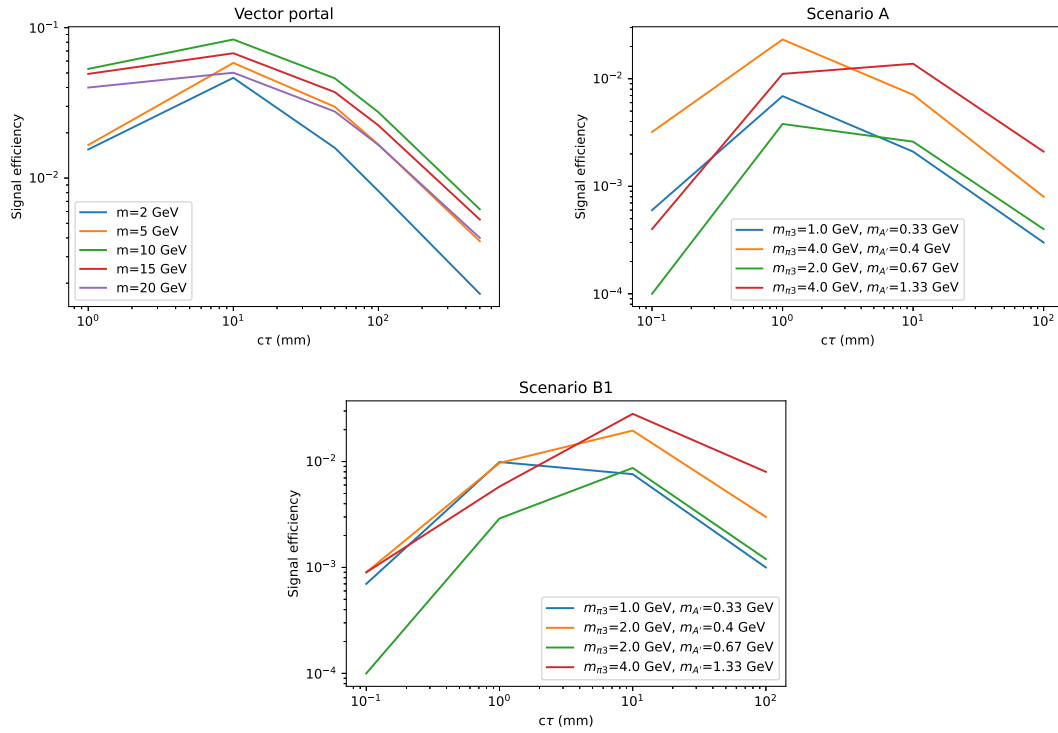


Figure E.1: Selection efficiency against the mean proper lifetime for benchmark signal models.

Appendix F

Impact plots

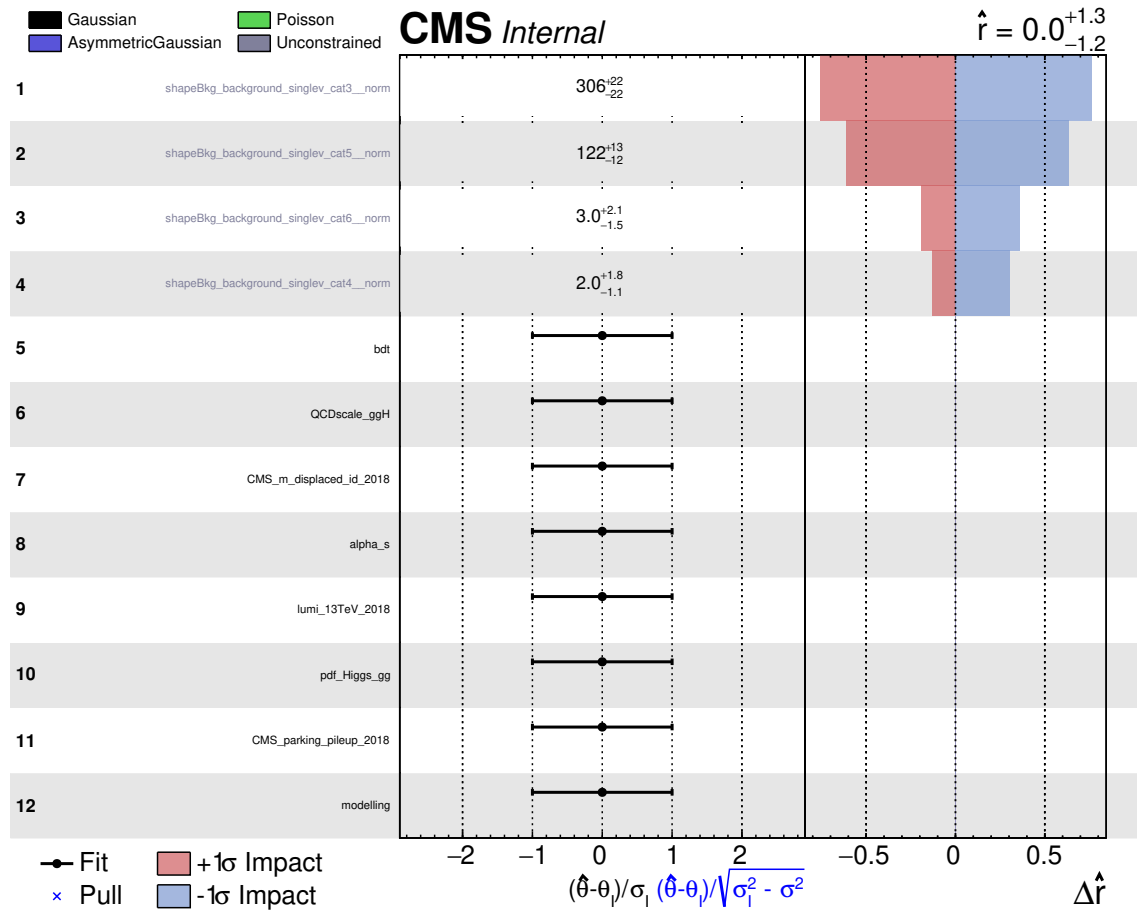


Figure F.1: Impact plot made using toys, without signal injection. The plot shows the fits of the nuisance parameters, and the impacts of the nuisance parameters on the signal strength. The scenario A model point with $m_{\pi 3} = 4$ GeV, $m_{A'} = 0.4$ GeV, $c\tau = 60$ mm is used for the study.

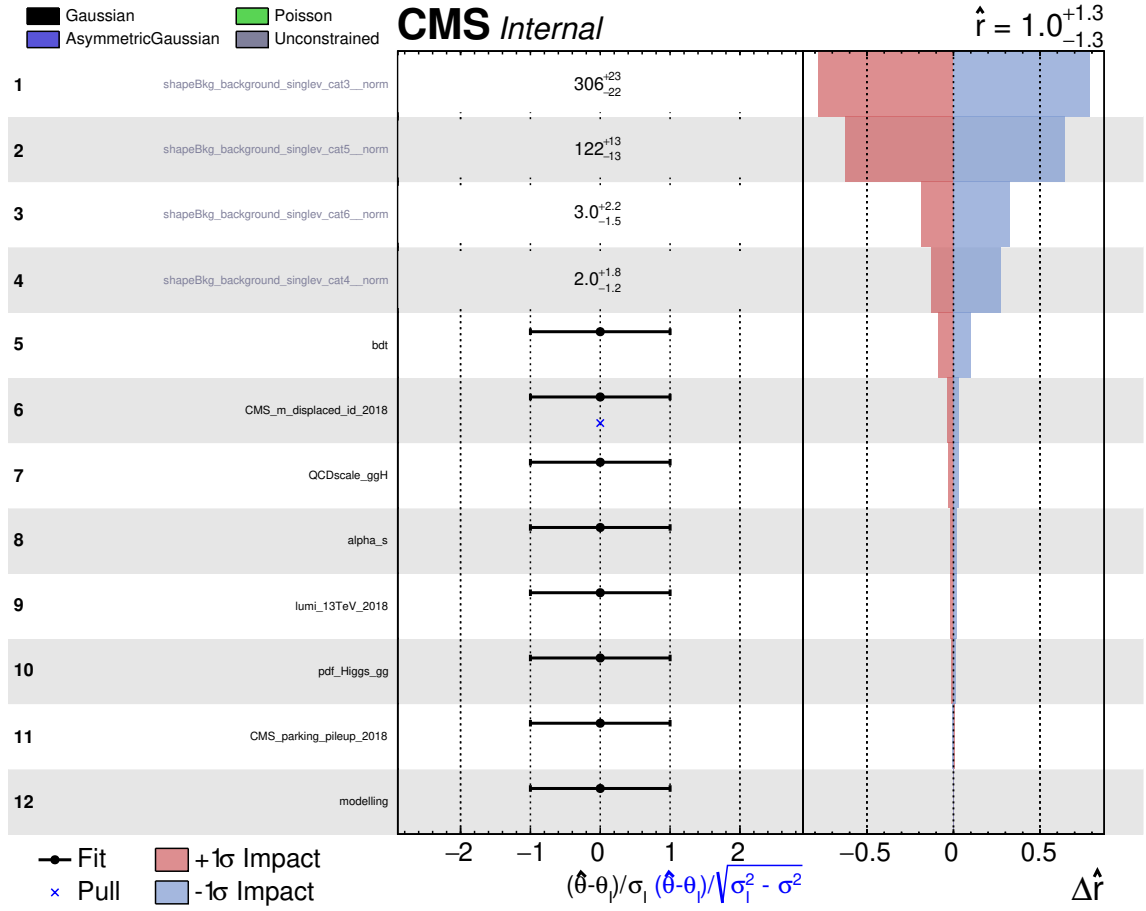


Figure F.2: Impact plot made using toys, with signal injection. The plot shows the fits of the nuisance parameters, and the impacts of the nuisance parameters on the signal strength. The scenario A model point with $m_{\pi 3} = 4$ GeV, $m_{A'} = 0.4$ GeV, $c\tau = 60$ mm is used for the study.

Appendix G

Dimuon mass distributions for scenario A and scenario B1

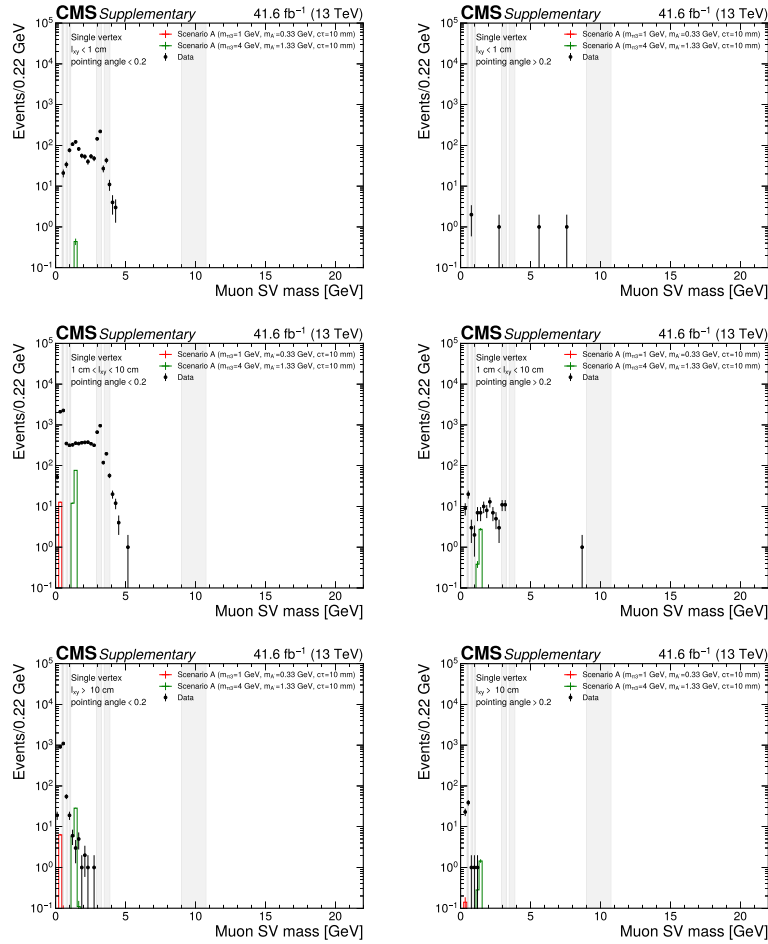


Figure G.1: Dimuon mass distributions in the single vertex categories for data and two benchmark signal models in scenario A. The shaded regions indicate mass regions of known SM resonances, which are masked in the search.

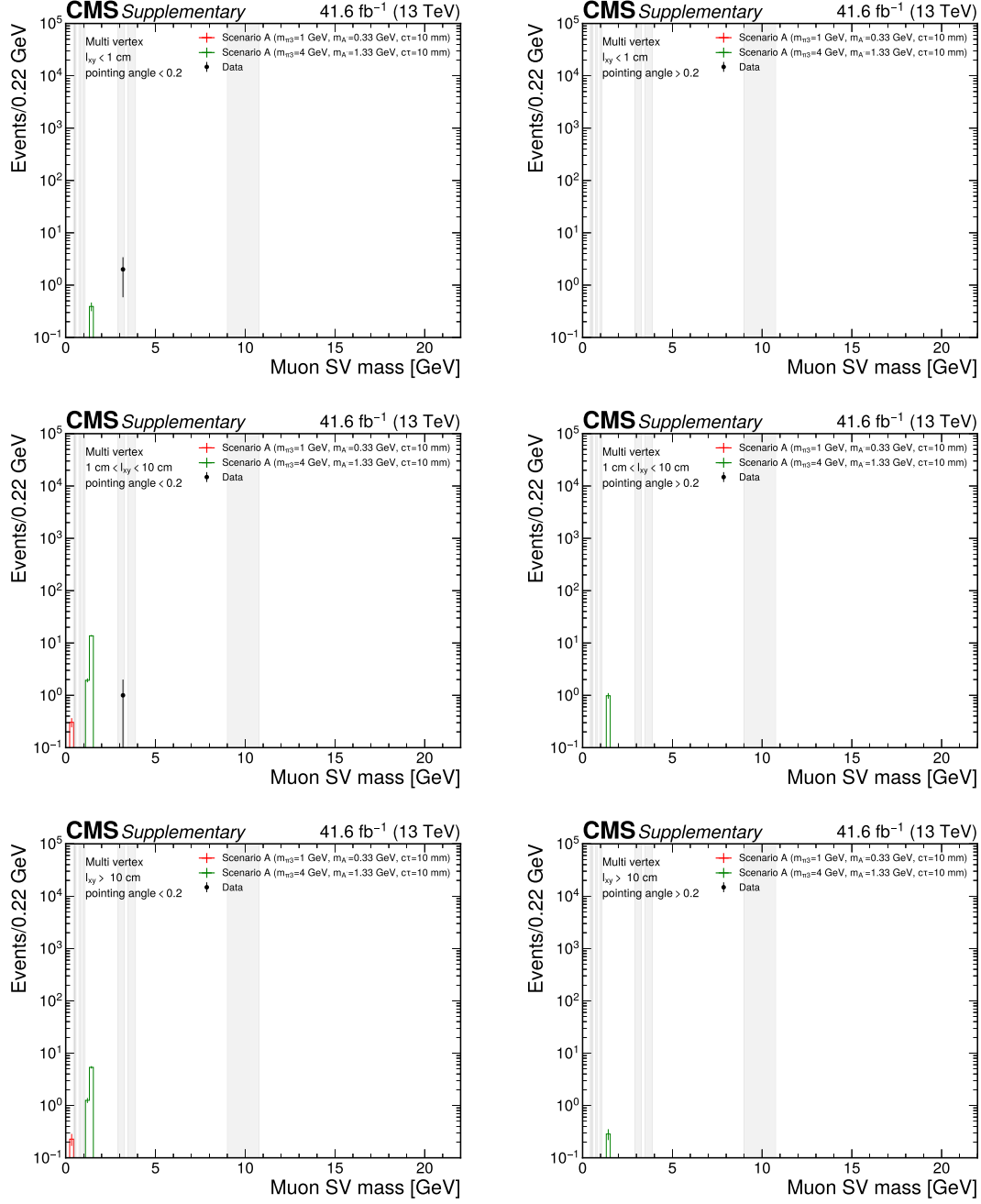


Figure G.2: Dimuon mass distributions in the multi vertex categories for data and two benchmark signal models in scenario A. The shaded regions indicate mass regions of known SM resonances, which are masked in the search.

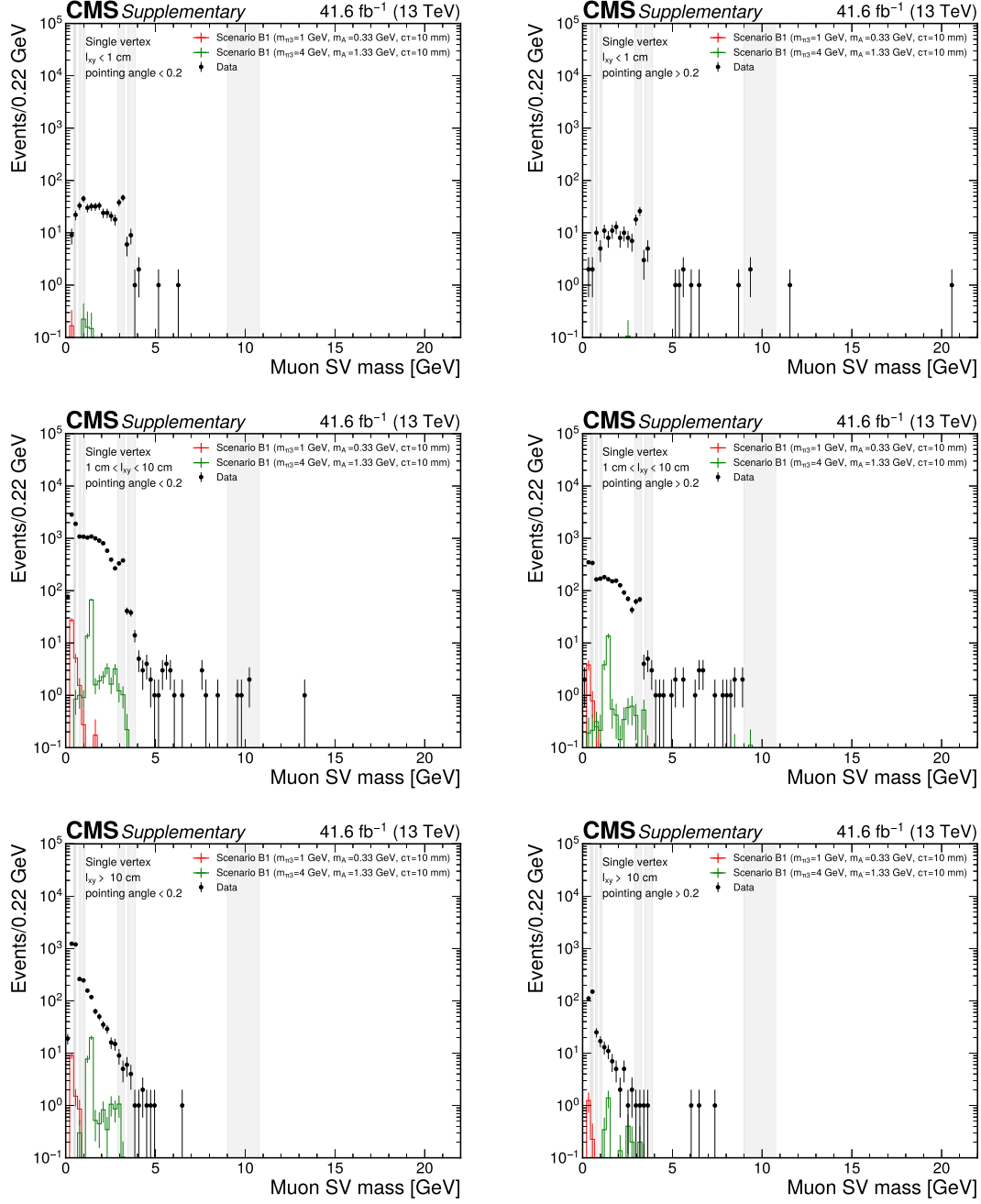


Figure G.3: Dimuon mass distributions in the single vertex categories for data and two benchmark signal models in scenario B1. The shaded regions indicate mass regions of known SM resonances, which are masked in the search.

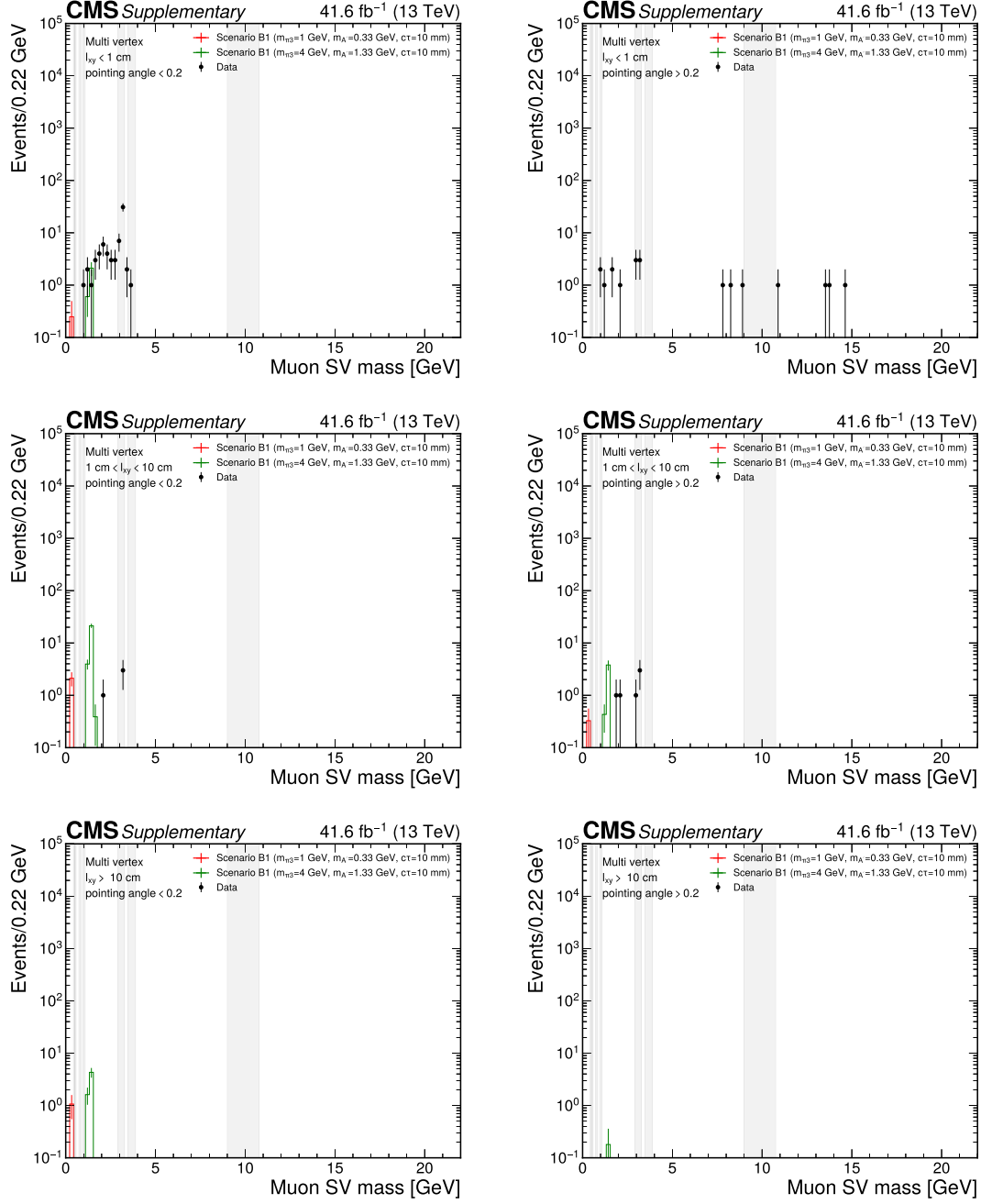


Figure G.4: Dimuon mass distributions in the multi vertex categories for data and two benchmark signal models in scenario B1. The shaded regions indicate mass regions of known SM resonances, which are masked in the search.