

SEARCH FOR THE HIGGS BOSON DECAYING TO A PAIR
OF MUONS WITH THE CMS EXPERIMENT AT THE LARGE
HADRON COLLIDER

by

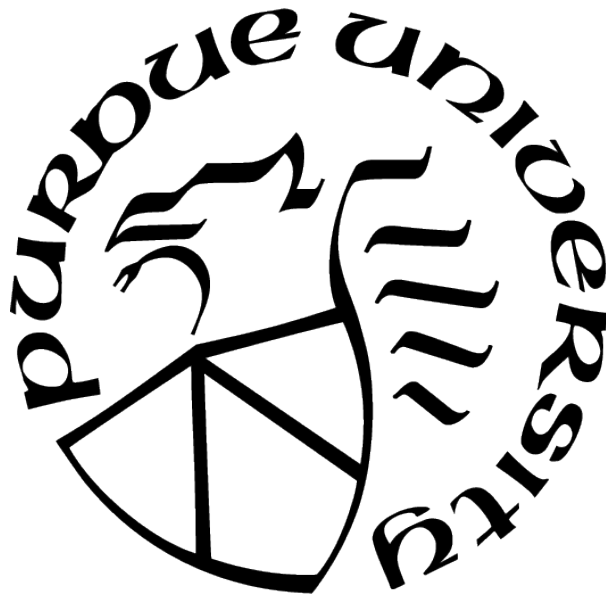
Dmitry Kondratyev

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Physics and Astronomy

West Lafayette, Indiana

December 2022



**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL**

Dr. Norbert Neumeister, Chair

Department of Physics and Astronomy

Dr. Andreas Jung

Department of Physics and Astronomy

Dr. Martin Kruczenski

Department of Physics and Astronomy

Dr. Yexiang Xue

Department of Computer Science

Approved by:

Dr. Gabor Csathy

Head of the Department of Physics and Astronomy

ACKNOWLEDGMENTS

The progress and achievements described in this thesis would not be possible without all the support and assistance I received from numerous people along my academic journey.

First and foremost, I would like to thank my scientific advisor, professor Norbert Neumeister. His guidance from the early days of my graduate studies helped me navigate through the ups and downs of scientific research, while the high standards and expectations, as well as openness to new ideas, have shaped me as an independent researcher.

I am also thankful to Dr. Andreas Jung, Dr. Martin Kruczenski, and Dr. Yexiang Xue for agreeing to join my Ph.D. committee and engaging in a fruitful discussion following my Ph.D. defense.

A significant part of my progress was due to immense help from my senior colleagues at the CMS collaboration. Among them, the most impactful were the advice and expertise provided by my co-authors — Jan-Frederik Schulte, Arnab Purohit, and Stefan Piperov. Not less important were the ideas and insights born in conversations and arguments with fellow CMS students at Purdue — Jason Thieman, AJ Wildridge, Amandeep Singh, and many others. I also cannot underestimate the help of my friends at CERN, Santeri Laurila and Jaana Heikkila, who were of great support during my time in Geneva.

Lastly, I am thankful to my friends and family in Russia, the USA, and around the world.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	x
ABSTRACT	xvi
INTRODUCTION	1
1 THEORETICAL OVERVIEW	3
1.1 The Standard Model of particle physics	3
1.1.1 The fundamental fermions	5
1.1.2 Gauge fields	7
Strong interaction	7
Electroweak interaction	8
Couplings to fermions	9
Kinetic terms	9
1.2 The Higgs sector of the Standard Model	10
1.2.1 The Higgs mechanism	10
1.2.2 Couplings to gauge bosons	12
1.2.3 Couplings to fermions	13
1.2.4 Limitations and prospects of the Higgs sector	14
1.3 Higgs physics at the LHC	16
1.3.1 Higgs boson production in proton-proton collisions	16
1.3.2 Higgs boson decay modes	19
1.3.3 Natural width of the Higgs boson	22
2 EXPERIMENTAL SETUP	23
2.1 The Large Hadron Collider	23
2.1.1 The CERN accelerator complex	25
2.1.2 Instantaneous and integrated luminosity	27

2.1.3	Pile-up	28
2.1.4	The LHC timeline	29
2.2	The CMS Experiment at the LHC	30
2.2.1	The coordinate system	31
2.2.2	The magnet	32
2.2.3	The tracker	33
2.2.4	The electromagnetic calorimeter	35
2.2.5	The hadron calorimeter	36
2.2.6	The muon system	38
2.2.7	The trigger system	40
2.2.8	Event reconstruction	41
3	CMS MUON HIGH LEVEL TRIGGER	44
3.1	Muon tracking, reconstruction and triggering algorithms	45
3.2	L3 muon reconstruction in Run 2	56
3.2.1	Outside-in seeding in Run 2	57
3.3	Outside-in seeding developments for Run 3	58
3.3.1	Changes to the ad hoc features remaining from the Run 2 strategy	59
3.3.2	Hitbased doublet seeds	60
3.4	DNN-based seeding strategy optimization	63
3.4.1	Problem definition	63
3.4.2	A new optimization approach	64
3.4.3	Input features	65
3.4.4	DNN architecture and training settings	66
3.4.5	Performance	71
4	FIRST EVIDENCE OF THE $H \rightarrow \mu\mu$ DECAY	75
4.1	Analysis strategy	76
4.2	Datasets	78
4.3	Event selection and categorization	79
4.4	Corrections to data and MC	80

4.5	Systematic uncertainties	83
4.5.1	Theoretical uncertainties	83
4.5.2	Experimental uncertainties	84
4.6	VBF category	85
4.6.1	Event selection and Data/MC agreement	85
4.6.2	DNN classifier	90
4.6.3	Fit procedure	92
4.6.4	Comparison between implementations	98
4.7	ggH category	101
4.7.1	BDT categorization	101
4.7.2	Signal modeling	103
4.7.3	Background modeling	105
4.7.4	Fit procedure	107
4.8	$t\bar{t}H$ category	108
4.9	WH and ZH categories	111
4.10	Combined results	115
5	NEW DEVELOPMENTS FOR THE $H \rightarrow \mu\mu$ ANALYSIS STRATEGY	121
5.1	Studies for the MC template-based strategy	121
5.1.1	Dimuon mass resolution	122
5.1.2	Jet pair selection	126
5.1.3	Event categorization in 2-jet phase space	130
5.1.4	Breakdown of signal sensitivity in VBF phase space	133
5.1.5	Template-based strategy in 1-jet phase space	134
5.2	MVA methods in the VBF phase space	137
5.2.1	Baseline DNN setup	137
	DNN architecture and training settings	137
	Importance of input features	139
5.2.2	Training datasets	142
5.2.3	Loss functions	144

5.3	Summary	145
6	PROSPECTS FOR THE PRECISE MEASUREMENT OF THE HIGGS BOSON PROPERTIES IN THE $H \rightarrow \mu\mu$ DECAY CHANNEL AT THE HL-LHC	148
6.1	Introduction	148
6.2	The CMS Phase-2 detector	150
6.3	Extrapolation of the Run 2 $H \rightarrow \mu\mu$ results to the HL-LHC	151
6.3.1	Scaling of cross sections	154
6.3.2	Detector acceptance	155
6.3.3	Dimuon mass resolution	156
	ggH category	157
	VBF category	157
6.3.4	Uncertainty scenarios	160
6.4	Results	162
	SUMMARY	166
A	DATASETS	168
B	MACHINE LEARNING METHODS IN PARTICLE PHYSICS	173
B.1	Deep Neural Networks (DNN)	173
B.1.1	DNN as a method to approximate a likelihood ratio	174
B.1.2	DNN as a tool to exploit correlations between parameter spaces	176
B.1.3	Architecture	176
B.1.4	Loss functions	178
B.1.5	Cross-validation	179
B.1.6	Criterion for the best model choice	180
B.2	Boosted Decision Trees (BDT)	182
	BIBLIOGRAPHY	184

LIST OF TABLES

1.1	Masses of the SM fermions and experimental limitations on the neutrino masses [23]. The neutrino mass states (ν_1, ν_2, ν_3) do not correspond to the flavor states, and the absolute values of their masses have not been measured. The only known quantities are differences between the neutrino mass states, but their order is unknown.	6
1.2	Main production channels of the Higgs boson in proton-proton collisions at a center-of-mass energy of 13 TeV.	17
1.3	Branching ratios of the Higgs boson decay modes for $M_H = 125$ GeV [42].	20
2.1	Nominal parameters of the LHC accelerator.	27
3.1	Error scale factors used in Run 2 [108].	55
4.1	Experimental uncertainties in the $H \rightarrow \mu\mu$ event (sub-)categories.	84
4.2	Expected and observed significance for two implementations of the VBF $H \rightarrow \mu\mu$ analysis.	99
4.3	Observed and expected significances for the incompatibility with the background-only hypothesis for $m_H = 125.38$ GeV and the corresponding 95% CL upper limits on μ (in the absence of $H \rightarrow \mu\mu$ decays) for each production category, as well as for the 13 TeV and the 7+8+13 TeV combined fits.	118
5.1	The distribution of the events in the total Run 2 VBF $H \rightarrow \mu\mu$ dataset by the number of selected reconstructed jets, as well as by the number of those jets matched with a generated jet.	127
5.2	The distribution of the events in the in the total Run 2 VBF $H \rightarrow \mu\mu$ dataset by the number of selected reconstructed jets, as well as by the number of those jets matched with a generated jet.	127
5.3	Accuracy of jet pair selection methods evaluated using events with at least 3 reconstructed jets, of which at least two are matched to generated jets, but not all of them.	128
5.4	Contribution of different components of the $H \rightarrow \mu\mu$ signal in the VBF phase space.	133
6.1	Previous projections made for $L = 3000 \text{ fb}^{-1}$	150
6.2	Event and object selections in Run 2 and HL-LHC	153
6.3	Total change in event yields between Run 2 and HL-LHC analyses. The yields are quoted for $L = 60 \text{ fb}^{-1}$	153
6.4	Cross sections at 13 TeV and 14 TeV.	154

6.5	Effect of updated detector acceptance: total event yields for Run 2 and HL-LHC analyses with all cross sections scaled to their 13 TeV values. The yields are quoted for $L = 60 \text{ fb}^{-1}$	155
6.6	Relative change of the inclusive event yields due to changes to cross sections and detector acceptance.	156
6.7	Comparison between different uncertainty scaling scenarios.	162
6.8	Uncertainty on the signal strength modifier $r^{\mu\mu}$ at $L = 3000 \text{ fb}^{-1}$	163
6.9	Uncertainty on the coupling modifier κ_μ at $L = 3000 \text{ fb}^{-1}$	163
A.1	Signal simulated samples for the 2016 data taking era.	168
A.2	Signal simulated samples for the 2017 data taking era.	169
A.3	Signal simulated samples for the 2018 data taking era.	169
A.4	Background simulation samples for the 2016 data taking era.	170
A.5	Background simulation samples for the 2017 data taking era.	171
A.6	Background simulation samples for the 2018 data taking era.	172

LIST OF FIGURES

1.1	Schematic representation [19] of the particles and interactions in the Standard Model. In addition to the particles depicted here, the SM includes an antiparticle for each of the fermions.	4
1.2	The shape of the Higgs potential for the case of $\mu^2 > 0$ [30].	11
1.3	Cross sections of Higgs boson production channels in proton-proton collisions [42].	17
1.4	Branching ratios of the Higgs boson decay modes [42].	19
2.1	Overall view of the LHC [77].	24
2.2	Accelerator complex at CERN (as of 2013) [80].	26
2.3	Integrated luminosity recorded by the CMS detector during LHC Runs 1 and 2 [81].	28
2.4	Distributions of the number of pile-up interactions in each data-taking period [81].	28
2.5	LHC timeline for Run 3 and HL-LHC (Runs 4 and 5 [83]).	29
2.6	A view of the CMS detector and its subsystems [74].	30
2.7	The map of the magnetic field strength (left), and the field lines (right) in the CMS detector [84].	32
2.8	Original layout of the CMS tracker [87].	33
2.9	Geometry of the pixel detector before and after the Phase-1 upgrade [88].	34
2.10	Schematic view of the CMS ECAL [91].	36
2.11	Schematic view of the CMS HCAL [93].	37
2.12	Schematic view of the CMS muon system [97].	39
2.13	The transverse momentum resolution of muons: RMS of q/p_T residuals distribution as a function of p_T measured using cosmic rays [97].	43
3.1	L2/L1 efficiencies in the Run 2 data [106].	51
3.2	Left: a <i>hitless seed</i> is created by propagating the L2 trajectory to a detector layer. Center: a <i>hitbased seed</i> is created by updating the propagated trajectory with the most compatible hit. Right: a <i>hitbased doublet</i> is created by updating the trajectory with two hits on consecutive layers.	52
3.3	L3(OI)/L2 efficiency for trajectory building seeded with a single HB or single HL seed.	53
3.4	L3(OI)/L2 efficiency as a function of pile-up. The increase in efficiency from additional seeds is correlated with the number of pile-up vertices.	54

3.5	L3(OI)/L2 efficiency for trajectory building seeded with various numbers of hitless seeds.	55
3.6	L3/L1 efficiencies in Run 2 [106].	57
3.7	Total number of seeds created in the Run 2 strategy, as a function of η	59
3.8	Proposed muon HLT design for Run 3.	59
3.9	Left: a regular HB seed constructed from the most compatible hit which can be from pile-up. Center: a seed is not generated if there are no compatible hits after update. Right: a HB-doublet seed is generated if there are compatible hits on two layers.	61
3.10	Efficiency with a single HB-doublet seed is similar to that of two regular HB seeds.	62
3.11	Processing time of seed generation and trajectory building for regular and doublet-like HB seeds.	62
3.12	DNN classifier used to predict the seeding strategy with highest probability of success.	64
3.13	Correlations between the DNN input features.	67
3.14	Distribution of some L2 parameters for events passing and failing L3 reconstruction seeded by a single HL(IP) seed. Variables labeled “err” are square roots of diagonal elements of covariance matrices corresponding to the L2 trajectories, i.e. uncertainties on global trajectory parameters.	69
3.15	Distribution of some L2 parameters for events passing L3 reconstruction seeded by different types of seeds. The DNN classifier is trained to learn the differences between these distributions.	70
3.16	Example of seeding strategies predicted by the DNN classifier trained to choose between 2-seed strategies without error scale factor optimization. In this example, the strategy with 1 HB-doublet and 1 HL(IP) seed is preferred for high-quality muons, while strategies with only hitless seeds are preferred for lower quality muons.	72
3.17	L3 efficiency of optimized strategies with fixed total number of seeds ranging from 1 to 7.	73
3.18	Training and validation loss for DNN selecting between 5-seed strategies, used in the barrel region (left), and for DNN selecting between 7-seed strategies, used in the overlap region (right).	73
3.19	L3 efficiency for the strategy optimized using DNN classifier, compared to the default Run 2 strategy. The efficiency is measured using a Drell–Yan dataset.	74
3.20	Seed generation (left) and trajectory building (right) time for the DNN classifier compared to the Run 2 approach. The timing is measured using a Drell–Yan dataset.	74

4.1	Feynman diagram of the $H \rightarrow \mu\mu$ decay.	75
4.2	Event categorization in the Run 2 $H \rightarrow \mu\mu$ analysis [1].	80
4.3	Distributions of dimuon p_T (top), leading muon p_T (middle) and subleading muon p_T (bottom) for 2016 (left), 2017 (center) and 2018 (right).	87
4.4	Distributions of p_T and η for the leading jet (top two rows), and for the subleading jet (bottom two rows) and subleading muon p_T (bottom) for 2016 (left), 2017 (center) and 2018 (right).	88
4.5	Distributions of dijet invariant mass (top), p_T -balance ratio (R_{p_T}) (middle) and logarithm of the Zeppenfeld variable z^* (bottom) for 2016 (left), 2017 (center) and 2018 (right).	89
4.6	Receiver operating characteristic (ROC) curve for the DNN used in the VBF category. Signal (background) efficiency is defined as signal (background) yield below a given DNN score cutoff, divided by the total signal (background) yield. The point (0,0) corresponds to the lowest DNN score, while (1,1) corresponds to the highest score. Blue markers correspond to the bin boundaries used in the Run 2 analysis.	91
4.7	Pre-fit (top row) and post-fit (bottom row) DNN score templates in VBF-SB region (left) and VBF-SR region (right) shown for the 2016 data-taking period.	94
4.8	Pre-fit (top row) and post-fit (bottom row) DNN score templates in VBF-SB region (left) and VBF-SR region (right) shown for the 2017 data-taking period.	95
4.9	Pre-fit (top row) and post-fit (bottom row) DNN score templates in VBF-SB region (left) and VBF-SR region (right) shown for the 2018 data-taking period.	96
4.10	Post-fit DNN score templates in VBF-SB region (left) and VBF-SR region (right), combined for all three data-taking periods.	97
4.11	Impacts of the most important systematic uncertainties on the signal strength fit in the VBF category.	98
4.12	Correlation of the DNN models evaluated using 2016 (top left), 2017 (top right), 2018 (bottom) datasets. The x-axis corresponds to the DNN score in the implementation illustrated in this thesis, while the y-axis shows the score distribution for the model used in Ref. [1]. The red lines are the edges of the DNN bins, while the blue cells denote the DNN bins that correspond to each other between the two implementations (the signal yield in these bins is synchronized by design).	100
4.13	Left: distribution of the BDT score in the ggH category and the subcategory boundaries. Right: comparison of signal mass resolution in a low-score and a high-score BDT subcategories [1].	103

4.14	Different signal models considered to parameterize the mass spectrum of the ggH $H \rightarrow \mu\mu$ signal. Top row, from left to right: sum of two Gaussians, Crystal Ball, sum of Crystal Ball and a Gaussian. Bottom row, from left to right: sum of three Gaussians, double-sided Crystal Ball, sum of double-sided Crystal Ball and a Gaussian.	104
4.15	Signal-plus-background fits in the ggH subcategories [1].	107
4.16	BDT score distributions and subcategory boundaries in $t\bar{t}H$ -hadronic (left) and $t\bar{t}H$ -leptonic (right) categories [1].	110
4.17	Signal-plus-background fits in the $t\bar{t}H$ subcategories [1].	111
4.18	BDT score distributions and subcategory boundaries in WH (left) and ZH (right) categories [1].	113
4.19	Signal-plus-background fits in the VH subcategories [1].	114
4.20	Left: observed local p -values as a function of m_H , extracted from the combined fit as well as from each individual production category, are shown. The solid markers indicate the mass points for which the observed p -values are computed. Right: the expected p -values are calculated using the background expectation obtained from the signal-plus-background fit and injecting a signal with $m_H = 125.38$ GeV and $\mu = 1$ [1].	116
4.21	Signal strength modifiers measured for $m_H = 125.38$ GeV in each production category (black points) are compared to the result of the combined fit (solid red line) and the SM expectation (dashed grey line) [1].	117
4.22	Observed (solid black) and expected (dashed black) local p -values as a function of m_H , extracted from the combined fit performed on data recorded at center-of-mass energies of 7, 8, and 13 TeV, are shown. The expected p -values are calculated using the background expectation obtained from the S+B fit and injecting a signal with $m_H = 125.38$ GeV and $\mu = 1$ [1].	119
4.23	The best fit estimates for the reduced coupling modifiers extracted for fermions and weak bosons from the resolved κ -framework compared to their corresponding prediction from the SM. The error bars represent 68% CL intervals for the measured parameters. In the lower panel, the ratios of the measured coupling modifiers values to their SM predictions are shown [1].	120
5.1	Correlation between the dimuon mass and the VBF DNN score for VBF $H \rightarrow \mu\mu$ events. The color represents the relative event-by-event mass resolution $\sigma_m/m_{\mu\mu}$	123
5.2	Relative per-event dimuon mass resolution distribution in 2016 (top left), 2017 (top right), 2018 (bottom).	124

5.3	Left: expected significance in the VBF channel as a function of the mass resolution improvement. The improvement is estimated by varying only the $m_{\mu\mu}$, σ_m , and $\sigma_m/m_{\mu\mu}$ inputs of the VBF DNN. Resolution improvement of 100% corresponds to $m_{\mu\mu} = 125$ GeV for all ggH and VBF signal events. Right: dimuon mass resolution for $H \rightarrow \mu\mu$ events as a function of the largest absolute pseudorapidity of the two muons.	125
5.4	ggH vs. VBF categorization efficiency in Run 2, compared to the new MVA-based method. Red: Run 2 efficiencies. Blue and orange: DNN setup with 3 input features. Green: DNN setup with 15 input features.	132
5.5	DNN score distribution in the VBF channel. Top left: 2016, top right: 2017, bottom: 2018.	135
5.6	DNN score distribution in 1-jet category for 2016, 2017 and 2018 data-taking periods.	136
5.7	DNN score distribution in the VBF channel for the baseline setup replicating the Run 2 sensitivity.	140
5.8	Relative importance of DNN input features estimated as their individual sensitivity, a loss of overall sensitivity if impact of a given feature is removed by shuffling its values between signal and background events, and an average gradient at the DNN input node corresponding to the given feature.	141
5.9	Left: significance evaluated on the validation fold for different VBF $H \rightarrow \mu\mu$ datasets and their combination. Right: significance for trainings with variable amount of events in <code>powheg_pythia_dipole</code> dataset (other VBF datasets are not used).	144
5.10	Evolution of the expected significance evaluated for the validation fold for different loss functions.	146
6.1	Pseudorapidity of the leading muon for ggH $H \rightarrow \mu\mu$ (left) and VBF $H \rightarrow \mu\mu$ (right) datasets after full event selection.	156
6.2	Left: Inclusive ggH $H \rightarrow \mu\mu$ dimuon mass spectrum in Run 2 and HL-LHC. Right: Full width at half maximum (FWHM) of the Higgs peak as a function of the largest $ \eta $ of the two muons.	157
6.3	Left: Dimuon mass spectrum of the VBF $H \rightarrow \mu\mu$ signal in Run 2 and at HL-LHC. Right: Reweighting factors for estimation of the effect of mass resolution improvement in the VBF channel.	158
6.4	Original and reweighted according to the updated mass resolution DNN templates in VBF category for 2016 (top left), 2017 (top right) and 2018 (bottom) data-taking periods.	159
6.5	Nominal and varied DNN shapes in the original Run 2 datacards (blue and orange), and after mass-reweighting (green and red).	160

6.6	Uncertainty on signal strength modifier $r^{\mu\mu}$ (left) and κ_μ (right).	164
6.7	Effect of improved dimuon mass resolution in ggH channel for S2 uncertainty scenario. Left: Uncertainty on signal strength modifier $r^{\mu\mu}$. Right: Uncertainty on coupling modifier κ_μ .	164
6.8	Left: Extrapolation of the expected significance for S1 and S2 uncertainty scenarios. Right: Effect of the improved dimuon mass resolution on expected significance in ggH channel for S2 uncertainty scenario.	165
B.1	An example of a fully-connected DNN architecture.	177
B.2	Splitting into training, validation and evaluation dataset in the k -fold procedure with four folds.	179
B.3	Left: evolution of the training and validation losses, as well as the expected significance computed for the validation dataset (25% events). Right: validation loss and significance for three trainings with identical settings.	182
B.4	Graphical representation of an example decision tree.	183

ABSTRACT

The CERN Large Hadron Collider (LHC) offers a unique opportunity to test the Standard Model of particle physics. The Standard Model predicts the existence of a Higgs boson and provides accurate estimates for the strength of the interactions of the Higgs boson with other particles. After the discovery of the Higgs boson, the measurement of its properties, such as its couplings to other particles, is of paramount importance. The projects described in this thesis explore different aspects of one of such measurements — the search for the Higgs boson decay into a pair of muons ($H \rightarrow \mu\mu$), conducted by the CMS experiment at the LHC. This decay plays an important role in elementary particle physics, as it provides a direct way to measure the coupling of the Higgs boson to the muon. The first evidence of the $H \rightarrow \mu\mu$ decay was reported in 2020 as a result of an elaborate statistical analysis of the dataset collected by the CMS experiment during Run 2 of the LHC (2016–2018). The observed (expected) upper limit on the signal strength modifier for this decay at 95% confidence level was found to be 1.93 (0.81), constituting the most precise measurement to date. The details of this analysis, along with studies to establish possible directions for the development of the next iteration of the $H \rightarrow \mu\mu$ analysis using Run 3 data, are discussed in this thesis. In addition, a novel machine learning-based algorithm for the muon high level trigger is presented, which ultimately improves the data-taking efficiency of the CMS experiment, and hence, helps to increase the sensitivity of future $H \rightarrow \mu\mu$ searches. Finally, projections of the $H \rightarrow \mu\mu$ search sensitivity to the data-taking conditions at the High-Luminosity Large Hadron Collider are presented, estimating the achievable precision for future measurements of the Higgs boson properties.

INTRODUCTION

Particle physics pursues one of the most ambitious goals among scientific disciplines — to uncover the fundamental laws of the universe by studying phenomena occurring at subatomic scales. Unsurprisingly, such an ambition requires highly sophisticated technology and a great deal of ingenuity. It is not a coincidence that the largest scientific instrument on the planet — the Large Hadron Collider (LHC) — is a particle accelerator.

Nowadays, the discoveries made with modern large-scale experiments are not accomplished by individual research groups, but through the combined efforts of large international collaborations. I have had the privilege and pleasure to participate in such an experiment — the Compact Muon Solenoid (CMS) experiment at the LHC — where I conducted research and contributed to the experiment’s operations.

The focus of my research as a part of the CMS collaboration was on the search for the Higgs boson decaying into a pair of muons ($H \rightarrow \mu\mu$). I joined the $H \rightarrow \mu\mu$ analysis team in 2018, and spent the following two years getting acquainted with the research methods, and investigating potential improvements of the analysis. My main contributions during that period include extensive studies of multivariate methods and event categorization approaches, improvements to the $H \rightarrow \mu\mu$ signal parameterization, and a fully independent implementation of one of the analysis categories. An elaborate analysis using the full LHC Run 2 dataset (2016–2018) concluded in 2019, and lead to the first evidence of the $H \rightarrow \mu\mu$ decay [1, 2], which became one of the most important experimental results in Higgs physics.

In 2020, I focused on the computational aspects of the $H \rightarrow \mu\mu$ analysis, and more generally, on new approaches for data analysis in high energy physics. Using state-of-the art methods of array programming and distributed computing, I designed a software framework [3], which allows to reduce the $H \rightarrow \mu\mu$ analysis throughput time by a factor of 100, and hence significantly accelerates the rate at which new analysis ideas can be implemented and tested [4].

In 2021, I made a technical contribution to the CMS experiment by developing a novel machine learning-based algorithm for the muon reconstruction in the CMS high level trig-

ger [5, 6]. The new algorithm has been deployed by the CMS experiment for Run 3 of the LHC and has been used in data-taking since July 2022.

I have also contributed to the Snowmass 2022 initiative [7, 8], which aims to make a comprehensive projection for various directions of particle physics research, and to estimate the results achievable in the future, to the best of our current knowledge. For this project, I prepared a thorough extrapolation of the $H \rightarrow \mu\mu$ analysis sensitivity for the data-taking conditions at the High-Luminosity Large Hadron Collider (HL-LHC), which will operate in the 2030's–40's [9, 10].

My work in the final year of the Ph.D. program focused on the preparation of the next iteration of the $H \rightarrow \mu\mu$ analysis, which will be performed with data taken during the Run 3 of the LHC (2022–2025). I have identified the most important factors that defined and limited the analysis performance in the Run 2 $H \rightarrow \mu\mu$ search, and suggested possible directions of improvement, accompanying them with preliminary studies [11]. I have also participated in the CMS detector operations, controlling the quality of data acquisition during the first months of the Run 3 data-taking.

The most important studies and experimental results obtained during the course of my Ph.D. research are described in this thesis, along with the necessary theoretical and experimental context. The dissertation is structured as follows. Chapter 1 gives an overview of the Standard Model of particle physics, with a focus on the Higgs mechanism of the electroweak symmetry breaking. The Higgs boson production and decay modes accessible at the LHC are also described. Chapter 2 summarizes the design and specifications of the LHC and the CMS detector. Chapter 3 discusses the basic concepts of the muon reconstruction in the CMS experiment, in particular, in the trigger system. Improvements to the muon trigger algorithms are described, including a novel approach to seeding optimization based on machine learning. Chapter 4 describes in detail the search for the $H \rightarrow \mu\mu$ decay with the LHC Run 2 data, which lead to the first evidence of this decay. Chapter 5 expands on this subject, analyzing the most impactful factors affecting the $H \rightarrow \mu\mu$ analysis performance, and suggesting the directions of the analysis development. Finally, Chapter 6 features a study that extrapolates the Run 2 analysis sensitivity to the future data-taking conditions at the High-Luminosity LHC.

1. THEORETICAL OVERVIEW

The ultimate objective of physics research is to devise a theory that describes all observed phenomena, and successfully predicts those which have not been observed yet.

Rather than developing a single explanation for everything, scientific theories focus on describing the phenomena at a certain scale, or in a certain domain. Quantum mechanics, fluid dynamics, astrophysics describe the same reality, but from different perspectives and using different tools.

An important property of a theory is its agreement with other theories at the boundaries of their domains (e.g. quantum mechanics should not contradict classical mechanics at larger scales). A special place, therefore, belongs to the theory that aims to describe the world at the most fundamental scale. A theory like that can only be based on mathematical abstractions and general principles, therefore an experimental validation becomes of paramount importance: even the most elegant abstract theory is in danger of being speculative in the absence of an experimental evidence.

At the moment, the most comprehensive fundamental theory describes the world as a multitude of quantum fields interacting with each other following a set of probabilistic yet well defined rules. The vast majority of the observed phenomena can be formulated in terms of point-like objects, giving the name to the domain – *particle physics*. The theory itself is called the *Standard Model* [12–18].

1.1 The Standard Model of particle physics

The Standard Model of particle physics (SM) includes all known elementary particles of matter and three out of the four known fundamental interactions – *electromagnetic*, *weak*, and *strong*, but does not include gravity. The effects of gravity on the scales of particle physics are negligible, as the intensity of the gravitational interaction is 25 orders of magnitude smaller than that of the weak nuclear force, the weakest of the SM interactions.

The matter particles in the SM are represented by fermionic fields satisfying the Dirac equation, and the interactions are introduced as bosonic fields associated with gauge symmetries of the SM. Weak and electromagnetic forces are unified into a single *electroweak*

interaction, and the corresponding electroweak symmetry is spontaneously broken by the scalar Higgs field, as shown in Figure 1.1. This procedure is called *the Higgs mechanism*, and it provides the recipe for the SM particles to acquire their masses.

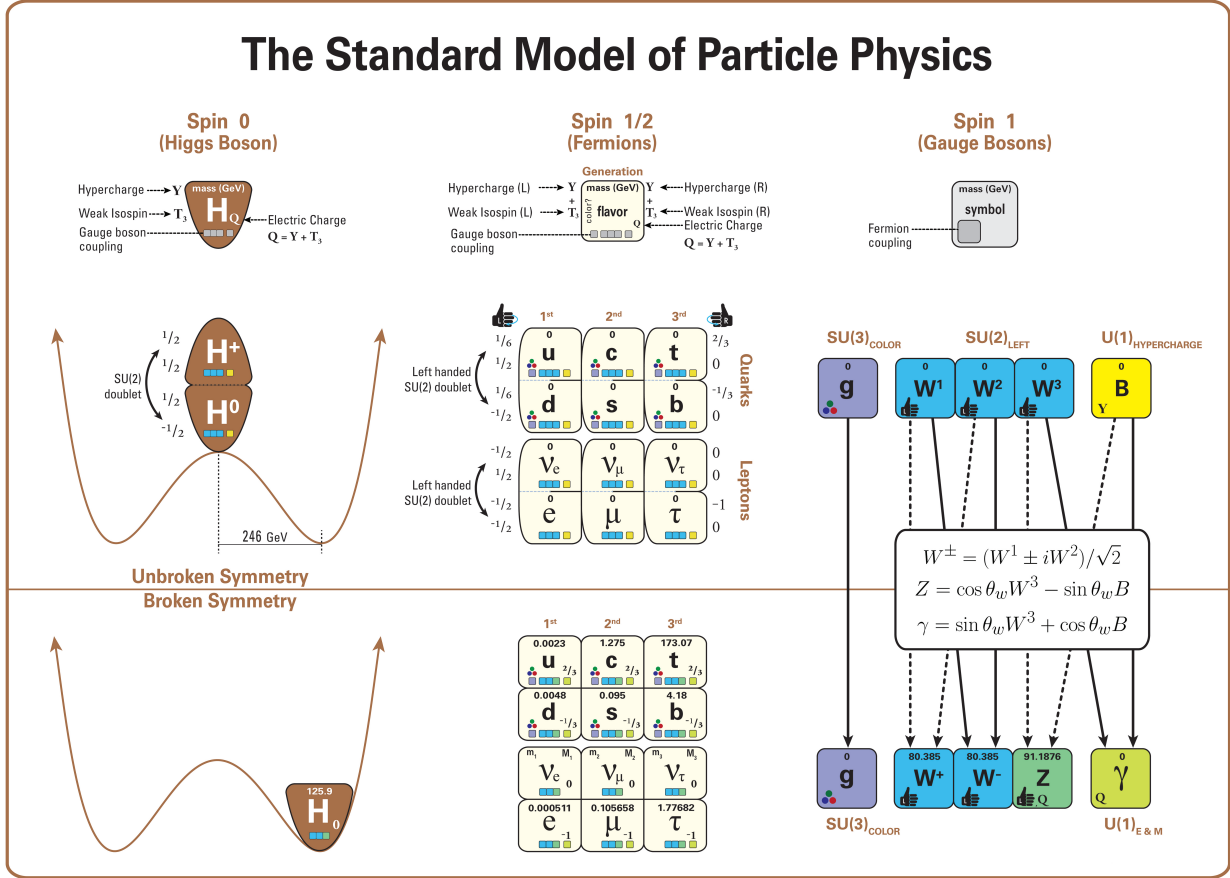


Figure 1.1. Schematic representation [19] of the particles and interactions in the Standard Model. In addition to the particles depicted here, the SM includes an antiparticle for each of the fermions.

The Standard Model can be formulated in terms of a *Lagrangian* — a single function of the fields and their derivatives, describing the dynamics and the degrees of freedom of a system. The evolution of the system can be derived using the stationary action principle, but a more practical application of the Lagrangian is calculation of probability amplitudes using Feynman diagrams.

The compact form of the SM Lagrangian density is shown in Eq. 1.1.

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\not{D}\psi + h.c. + \bar{\psi}_iy_{ij}\psi_j\phi + h.c. + |D_\mu\phi|^2 - V(\phi) \quad (1.1)$$

The first term describes the fundamental interactions between the SM particles, the second and the third terms define the couplings between the matter and the gauge fields (*h.c.* means *hermitian conjugate* of the preceding term); the fourth and the fifth terms describe the interactions between the fermions and the Higgs field, through which they acquire their masses, and the last two terms describe the Higgs field and its interactions with the gauge bosons. The meaning and the implications of the Lagrangian components are discussed in more detail in the following subsections.

1.1.1 The fundamental fermions

The fundamental particles of matter in the Standard Model are described by Dirac fields (ψ in Eq. 1.1) that correspond to spin- $\frac{1}{2}$ fermions. These fermions are classified according to their interactions with the gauge fields, which are discussed in detail in the next subsection.

The fundamental fermions that participate in the strong interaction are called *quarks*, while other fundamental fermions are called *leptons*. The quarks are further classified into *up-type* quarks, which carry electric charge $+2/3$, and *down-type* quarks, with charge $-1/3$. The leptons are classified into *charged leptons* with electric charge -1 , and electrically neutral *neutrinos*.

Spin and electric charge define the type of a fermion, but they do not uniquely identify a particle. The Standard Model features three particles for each fermion type, with differences only in their masses. The fermions are therefore arranged into three *generations*, each containing an up-type quark, a down-type quark, a charged lepton, and a neutrino. The SM fermions and their masses are listed in Table 1.1. The fermion masses are free parameters in the Standard Model, and their particular values are not derived from any underlying principles. The hierarchy of the three generations is defined by the masses of the fermions: the first generation contains the lightest fermions, while the third generation contains those with the highest masses. While the SM describes only three generations of fermions, it

does not prohibit additional generations; however, there has been no experimental evidence suggesting their existence.

The neutrinos are assumed to be massless in the original formulation of the Standard Model. However, an extension of the SM that includes massive neutrinos has found an experimental evidence [20, 21]. The exact values of the neutrino masses are yet to be determined, but it has been measured that they would not exceed 0.8 eV [22].

For each of the fundamental fermions, the SM also contains an *antiparticle* with the same mass and spin, but opposite physical charges (such as electric and color charge). The existence of the antifermions is required by the Dirac equation. The antiparticles of bosons are also included in the SM: the gluon, the photon, the Z boson and the Higgs boson are their own antiparticles, while the two W bosons (W^+ and W^-) are antiparticles to each other.

Table 1.1. Masses of the SM fermions and experimental limitations on the neutrino masses [23]. The neutrino mass states (ν_1, ν_2, ν_3) do not correspond to the flavor states, and the absolute values of their masses have not been measured. The only known quantities are differences between the neutrino mass states, but their order is unknown.

	1 st generation	2 nd generation	3 rd generation
Up-type quarks	u -quark 2.16 MeV	c -quark 1.27 GeV	t -quark 172.76 GeV
Down-type quarks	d -quark 4.67 MeV	s -quark 93 MeV	b -quark 4.18 GeV
Charged leptons	electron (e) 0.51 MeV	muon (μ) 105.6 MeV	τ -lepton 1776.86 MeV
Neutrinos	ν_e	ν_μ	ν_τ
	$\Delta m_{21}^2 = (7.53 \pm 0.18) \times 10^5 \text{ eV}^2$		
	$\Delta m_{32}^2 = (-2.524 \pm 0.034) \times 10^3 \text{ eV}^2$ (inverted order)		
	$\Delta m_{32}^2 = (2.453 \pm 0.033) \times 10^3 \text{ eV}^2$ (normal order)		
	$0.0 \text{ eV} < m_\nu < 0.8 \text{ eV}$ [22]		

1.1.2 Gauge fields

The fundamental interactions in the Standard Model arise from the requirement for the Lagrangian to be invariant with respect to certain local transformations. Such an invariance is referred to as *gauge symmetry*. The particular structure of the gauge symmetry in the Standard Model is

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y, \quad (1.2)$$

where c refers to the color charges of gluons and quarks, which act as generators of the $SU(3)$ group, L denotes the coupling of the weak gauge field to the fermions with only left-handed chirality, and Y refers to the weak hypercharge which is the generator of the $U(1)$ group.

Since the $U(1)$ group has one degree of freedom, and $SU(N)$ groups have $N^2 - 1$ degrees of freedom, the gauge symmetry group of the Standard Model has a total of $1 + 3 + 8 = 12$ degrees of freedom, and therefore twelve gauge bosons associated with it. All known gauge bosons in the SM have spin 1.

Strong interaction

The eight bosons associated with the $SU(3)_c$ symmetry group are called *gluons*, they mediate the strong nuclear interaction. The quarks under the $SU(3)_c$ symmetry transform as triplets, with each of the three states characterized by a different *color charge*.

Multiple quarks and/or anti-quarks can produce bound states via the strong interaction; it has been observed that only states with a net zero color charge are formed. The bound states consisting of a quark-antiquark pair are called *mesons*, while the three-quark states are called *baryons*¹. All particles participating in the strong interaction, including gluons, quarks, and their bound states, are collectively called *hadrons*.

An important property of the strong interaction is its *asymptotic freedom* [12, 24]: the strong coupling becomes weaker as the momentum transfer in the interaction increases. This allows to study the strong interaction at high-energy hadron colliders such as the LHC (see Chapter 2).

¹↑Notable examples of baryons are the proton (consisting of u, u, d quarks) and the neutron (consisting of u, d, d quarks).

Electroweak interaction

The $SU(2)$ and the $U(1)$ symmetries describe the two remaining interactions — weak and electromagnetic [13–15]. However, the phenomenology of the Standard Model is described more conveniently if these interactions are interpreted as a unified *electroweak* interaction with $SU(2)_L \otimes U(1)_Y$ symmetry group. The four generators of this groups correspond to three weak boson fields W_1 , W_2 , and W_3 , and the hypercharge boson field B . The charges associated with these fields are called *weak isospin* $T^{(3)}$ and *weak hypercharge* Y_W . The convenience of this representation is related to the (empirical) fact that the weak boson fields interact only with the left-handed chiral components of the fermion fields. Specifically, the left-handed components of quarks of the same generation form a doublet $\begin{pmatrix} u_L \\ d_L \end{pmatrix}$ under the action $SU(2)_L$, as well as the charged lepton and the corresponding neutrino: $\begin{pmatrix} e_L \\ \nu_L \end{pmatrix}$.

The W_1 , W_2 , W_3 , and B bosons are not observed directly. Instead, the boson fields are mixed, producing $W^\pm$, Z^0 bosons, and the photon. The observable states of the fields can be obtained from the W_1 , W_2 , W_3 , and B fields as follows:

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2) \quad (1.3)$$

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}, \quad (1.4)$$

under the condition $g \sin \theta_W = g' \cos \theta_W = e$, where g and g' are the coupling constants of the W_1 , W_2 , W_3 , and B fields, and e is the electric charge of an electron. θ_W is called *weak mixing angle*, it is established empirically and can be measured as $\cos \theta_W = \frac{m_W}{m_Z}$.

The weak isospin and the hypercharge are related to the electric charge Q of a particle via the Gell-Mann–Nishijima formula [25, 26]:

$$Q = T_3 + \frac{1}{2}Y_W. \quad (1.5)$$

The weak isospin and the hypercharge are conserved in all interactions, excluding the interactions with the Higgs field. The electric charge is conserved in all SM interactions.

Couplings to fermions

The interactions of the gauge fields with fermionic fields of matter are encoded by the covariant derivative, which appears in the second term of Eq. 1.1. The full form of the derivative can be written as

$$D_\mu = \partial_\mu + ig_s t^a G_\mu^a(x) + ig T^i W_\mu^i(x) + ig' Y B_\mu(x), \quad (1.6)$$

where ∂_μ is a regular space-time derivative, g_s , g , and g' are the coupling constants of the strong, weak, and hypercharge fields. $G_\mu^a(x)$, $W_\mu^i(x)$, $B_\mu(x)$ are the fields of the gluons and the electroweak bosons, and t^a , T^i , Y are the generators of the respective Lie groups.

Kinetic terms

The dynamics of the gauge fields are described by the kinetic term — the first term in Eq. 1.1. A kinetic term for a generic gauge field A_μ^a is obtained as a commutator of the corresponding covariant derivative:

$$F_{\mu\nu} = \frac{i}{g} [D_\mu, D_\nu] = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + ig f^{abc} A_\mu^b A_\nu^c. \quad (1.7)$$

The structure coefficient f^{abc} here is a constant, totally antisymmetric tensor that vanishes for the fields with abelian gauge groups, such as $U(1)$, but does not vanish for non-abelian groups, such as $SU(2)$ and $SU(3)$. Consequently, third and quartic vertices can be constructed for the weak and strong boson fields, but self-interactions of photons are not allowed.

1.2 The Higgs sector of the Standard Model

The gauge theory based on the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ symmetry group does not allow adding explicit mass terms for fermions or gauge bosons without breaking the gauge invariance. However, it has been observed that both the fundamental fermions and the electroweak bosons (excluding the photon) are massive. The solution to this problem is provided by the Higgs-Brout-Englert mechanism [16–18, 27–29], which introduces spontaneous electroweak symmetry breaking via an additional scalar field.

1.2.1 The Higgs mechanism

The Higgs field is introduced as a complex scalar field that is a doublet under the $SU(2)$ symmetry group:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi_0 \end{pmatrix}. \quad (1.8)$$

The corresponding boson — the *Higgs boson* — is the only fundamental particle in the Standard Model with spin 0. Its electric charge is zero, and its weak hypercharge is 1/2.

The potential of the Higgs field has the form

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad (1.9)$$

where λ and μ are real constants. The value of λ has to be positive in order to avoid an unphysical behavior of the potential for large values of ϕ . The shape of the potential depends on the value of μ^2 : if $\mu^2 < 0$, the potential has a trivial minimum at $\phi_{\text{ground}} = 0$. However, if $\mu^2 > 0$, the potential takes the *Mexican hat* shape, shown in Figure 1.2. For this shape, there is an infinite number of the states that minimize the potential, they are defined to satisfy the following requirement:

$$\phi^\dagger \phi = \frac{\mu^2}{2\lambda} = \frac{v^2}{2}, \quad (1.10)$$

where the parameter v is called the *vacuum expectation value*.

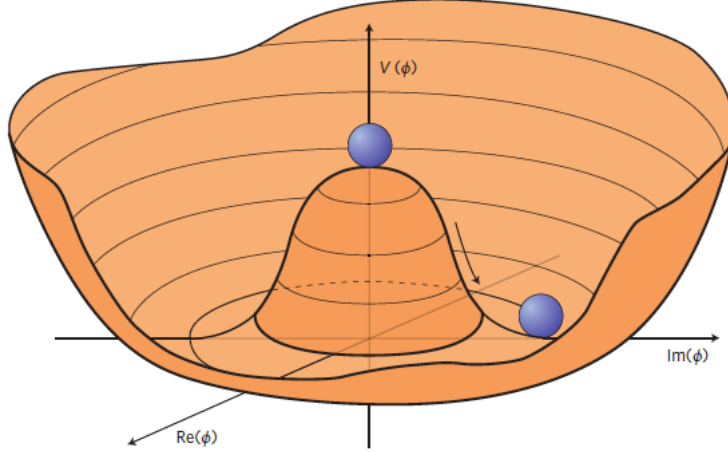


Figure 1.2. The shape of the Higgs potential for the case of $\mu^2 > 0$ [30].

An arbitrary state around the minimum can be expanded in the following form:

$$\phi_{\text{ground}}(x) = \frac{1}{\sqrt{2}} e^{-i\sigma_a \theta^a(x)/2} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \quad a = 1, 2, 3, \quad (1.11)$$

where $h(x)$ is a massive scalar field corresponding to the Higgs boson, $\sigma_{1,2,3}$ are the Pauli matrices, and $\theta_{1,2,3}(x)$ are massless scalar fields that appear as a consequence of the Goldstone theorem² [31]. The $\theta_{1,2,3}(x)$ fields can be absorbed through an $SU(2)_L$ transformation of the Higgs field without a loss of generality, leading to a simpler expression:

$$\phi'_{\text{ground}}(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (1.12)$$

The existence of the non-zero vacuum expectation value is a unique feature of the Higgs field: other fields in the SM do not exhibit the same behavior. In the lowest energy state, the Higgs field is forced to settle in one of the infinite number of minima, and the resulting Lagrangian does not satisfy the requirement of gauge invariance (specifically, that of the $SU(2)_L \otimes U(1)_Y$ group). This phenomenon is called *spontaneous electroweak symmetry breaking*, and it triggers the mechanism by which the bosons associated with the broken

²↑ The Goldstone theorem states that a spontaneous breaking of a continuous symmetry gives rise to as many massless bosons (*Goldstone bosons*) as there are broken generators of the symmetry.

symmetries become massive. This mechanism is discussed in more detail in the following subsection.

The vacuum state preserves invariance with respect to $U(1)_{em}$, therefore the electric charge remains conserved, and the photon does not acquire mass. The $SU(3)_c$ symmetry is also unaffected, therefore the gluons remain massless as well.

1.2.2 Couplings to gauge bosons

The kinetic term for the Higgs field has the following form:

$$K(\phi) = D_\mu \phi^\dagger D^\mu \phi \quad (1.13)$$

Taking the terms of the covariant derivative from Eq. 1.6 that correspond to the electroweak gauge bosons, in can be expanded to:

$$\frac{1}{8}g^2 (W_{1\mu}^2 + W_{2\mu}^2) (v + h(x))^2 + \frac{1}{8}g^2 (gW_{3\mu} - g'B_\mu)^2 (v + h(x))^2, \quad (1.14)$$

and then rewritten using the transformations outlined in Eq. 1.3 and 1.4:

$$\left(m_W W_\mu^+ W^{\mu-} + \frac{1}{2} m_Z Z^2 \right) \left(1 + 2 \frac{h(x)}{v} + \frac{h(x)^2}{v^2} \right), \quad (1.15)$$

where the constants $m_W = g^2 v^2 / 4$ and $m_Z = (g^2 + g'^2) v^2 / 4$ can be interpreted as masses of the $W^\pm$ and the Z bosons, respectively. The right factor in the equation above contains the terms involving $h(x)$ that represent the couplings between the Higgs boson and the weak gauge bosons, therefore the Higgs mechanism predicts triple and quartic vertices describing interactions between the Higgs field and the weak boson fields. The coupling constants of such interactions can be derived from the same equation:

$$\begin{aligned} g_{HVV} &= \frac{2m_V^2}{v}, \\ g_{HHVV} &= \frac{m_V^2}{v^2}. \end{aligned} \quad (1.16)$$

Thus, for gauge bosons, the coupling to the Higgs field is proportional to the mass of the boson squared.

1.2.3 Couplings to fermions

While the explicit mass terms for the SM fermions

$$m_f(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \quad (1.17)$$

cannot be included in the Lagrangian due to violation of the gauge symmetry, similar terms arise as a consequence of the interactions between the SM fermions and the Higgs field. These couplings are constructed as follows:

$$y_f \left(\bar{\psi}_L^{(u)}, \bar{\psi}_L^{(d)} \right) \phi \psi_R^{(d)} + h.c. = \left(\frac{y_f v}{\sqrt{2}} \bar{\psi}_L^{(d)} \psi_R^{(d)} + \frac{y_f}{\sqrt{2}} h \bar{\psi}_L^{(d)} \psi_R^{(d)} \right) + h.c. \quad (1.18)$$

$$y_f \left(\bar{\psi}_L^{(u)}, \bar{\psi}_L^{(d)} \right) \tilde{\phi} \psi_R^{(u)} + h.c. = \left(\frac{y_f v}{\sqrt{2}} \bar{\psi}_L^{(u)} \psi_R^{(u)} + \frac{y_f}{\sqrt{2}} h \bar{\psi}_L^{(u)} \psi_R^{(u)} \right) + h.c. \quad (1.19)$$

Eq. 1.18 describes the interactions of the Higgs field with down-type quarks and charged leptons, while Eq. 1.19 describes the same for the up-type quarks and the neutrinos. In the latter, the conjugate field $\tilde{\phi}(x) = i\sigma^2\phi(x)$ is used.

The first terms in the parentheses on the right side of Eq. 1.18 and 1.19 together with their hermitian conjugates have exactly the same form as Eq. 1.17, therefore they can be interpreted as mass terms for the fermionic fields, and the masses of the fermions are proportional to the corresponding Yukawa couplings:

$$m_f = \frac{y_f}{\sqrt{2}} v. \quad (1.20)$$

Therefore, the fermions acquire their masses as one of the consequences of their interaction with the scalar Higgs field, and their masses are proportional to the corresponding coupling constants. On the other hand, the masses of the fermions can be treated as free parameters of the Standard Model, in which case the conclusion can be reformulated as

follows: the coupling constants of the interactions of SM fermions with the Higgs field are proportional to the fermion masses.

Since the masses of most of the SM fermions have been measured long before the discovery of the Higgs boson, this property of the Standard Model allows to predict the branching fractions of the Higgs boson decays into fermions with high precision.

1.2.4 Limitations and prospects of the Higgs sector

The SM has been unexpectedly successful in describing with high precision the results of the grand majority of particle physics experiments. Yet despite of this success, the SM leaves important questions unanswered and its extrapolation to very high energies is unsatisfactory. In particular, the SM does not include gravity, dark matter [32–34] (the reason for the discrepancy between the cosmological observations of gravitational effects and the observed amount of detectable matter), dark energy [35, 36] (non-zero vacuum energy density), the matter–antimatter asymmetry in the universe [37], the strong CP problem [38], and neutrino masses [20, 22] (the SM predicts that, like photons, neutrinos should have no mass). Additionally, the SM does not provide an adequate explanation for some of its properties, such as the hierarchy of the particle masses, the number of generations of fermions, and the scale of the Higgs boson mass and the vacuum expectation value. The SM is usually formulated using 18 free parameters:

- 3 gauge couplings (g_s, g, g') describing the strong and electroweak forces;
- 3 mixing angles ($\theta_{12}, \theta_{23}, \theta_{13}$) of the Cabibbo-Kobayashi-Maskawa (CKM) matrix [39, 40] describing the mixing between quarks of different flavors, and the CP-violating phase δ_{13} ;
- the Higgs boson mass m_H and the Higgs potential vacuum expectation value v ;
- 9 Yukawa couplings describing the Higgs couplings to the massive fermionic fields of the SM (quarks and charged leptons).

Most of these parameters (the 9 Yukawa fermion mass terms, the Higgs mass, the Higgs potential vacuum expectation value, and the four CKM values) are related to the Higgs

sector, and inclusion of SM extensions that address some of the limitations mentioned above would further increase the number of parameters. These considerations motivate the research needed to be conducted to better understand the Higgs sector of the SM and possibly construct a theory capable of describing a wider range of phenomena with fewer free parameters.

The outstanding versatility of the Higgs mechanism is in its ability to explain some of the fundamental properties of the Standard Model via a relatively simple assumption about the existence of a field satisfying only a few basic requirements. These requirements include the existence of a non-trivial minimum of the potential, the invariance with respect to $U(1)_{em}$, and physical values at low energy scales. However, it might be that the approximation given by these assumptions does not fully describe the real structure of the Higgs field. For instance, the field could feature more than one set of local or global minima, which would lead to new symmetry breaking mechanisms and unexplored implications for the Standard Model.

In the SM, the Higgs boson is unique: it has zero spin, no electric charge and no strong force interaction. The spin and parity were measured through angular correlations between the particles it decayed to. The mass of the Higgs boson, as well as its self-interaction coupling are directly related to the μ and λ parameters of the Higgs potential as defined in Eq. 1.9. While the Higgs mass can be measured in a straightforward way by analyzing the Higgs boson decays, the measurement of the self-coupling requires studying processes with at least two Higgs bosons, which have significantly lower cross sections.

The Higgs mechanism by its design is expected to be connected to many of the theories that aim to describe the physics beyond the Standard Model (BSM), such as dark matter or supersymmetry [41]. The theories that predict the existence of additional massive particles must imply their interaction with the Higgs field.

Such interactions may manifest themselves via processes with similar final states as those predicted by the SM, leading to discrepancies between predicted and observed branching fractions. On the other hand, some of the BSM theories predict processes with final states prohibited by the Standard Model (e.g. those violating lepton flavor conservation), or states with undetectable decay products (e.g. dark matter particles), which would not directly

interfere with the SM predictions. Some theories also predict the existence of additional types of the Higgs boson, with different values of the Higgs mass and quantum numbers (electric charge and spin).

Thus, the Higgs sector serves as a window to new physics, hinting at the directions of possible expansions of the SM.

1.3 Higgs physics at the LHC

As shown in Sections 1.2.2 and 1.2.3, the strength of the interactions between the Higgs field and the SM particles is larger for more massive particles. Therefore, the dominant processes for the Higgs boson production must involve its interactions with the heaviest SM particles — t or b quarks, and Z or W bosons — which can only be produced in high-energy collisions.

Currently, the most powerful particle accelerator, and the only functioning accelerator capable of producing the Higgs boson, is the Large Hadron Collider (LHC), described in detail in the next chapter. At the LHC, the Higgs boson is produced in proton-proton collisions at center-of-mass energies up to 14 TeV. This section describes the most common Higgs boson production modes at the LHC, as well as its main decay channels.

1.3.1 Higgs boson production in proton-proton collisions

The cross sections of the Higgs production channels in proton-proton collisions depend on the Higgs boson mass, as well as the center-of-mass energy of the protons (see Figure 1.3). The fact that the Higgs mass happens to be $\mathcal{O}(100 \text{ GeV})$ is rather fortunate, as it makes the Higgs boson production experimentally accessible at the LHC. The growth of the cross sections with energy motivates the need for a high-energy collider such as the LHC, to produce Higgs bosons at a higher rate.

For $M_H = 125 \text{ GeV}$ and $\sqrt{s} = 13$ or 14 TeV , the dominating production channels are summarized below in decreasing order of cross section; their Feynman diagrams and cross sections are listed in Table 1.2.

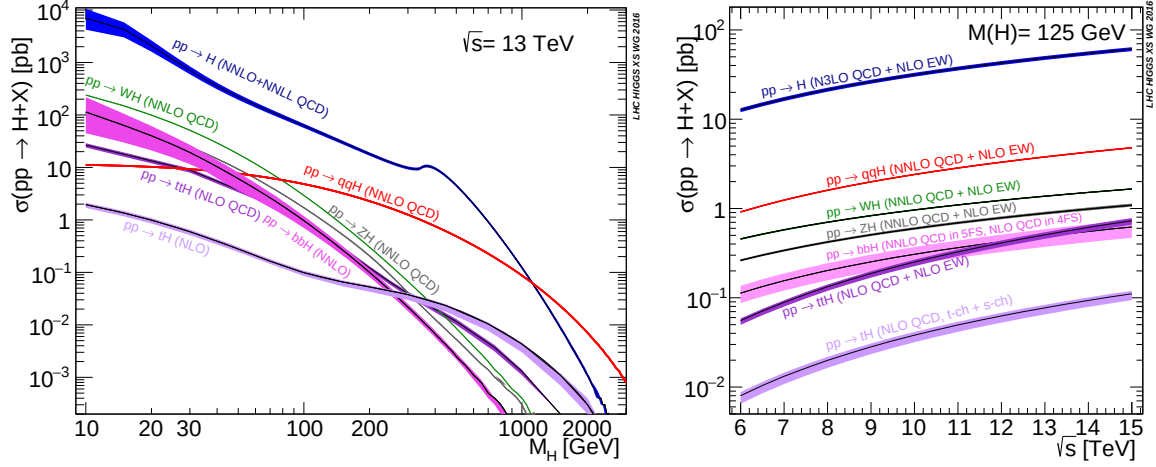


Figure 1.3. Cross sections of Higgs boson production channels in proton-proton collisions [42].

Table 1.2. Main production channels of the Higgs boson in proton-proton collisions at a center-of-mass energy of 13 TeV.

Production channel	Feynman diagram	Cross section [43]
ggH		48.58 pb [44] N3LO QCD ³ , NLO EW ⁴
VBF		3.748 pb NNLO QCD, NLO EW
VH		1.380 pb (WH) 0.870 pb (ZH) NNLO QCD, NLO EW
ttH		0.507 pb NLO QCD, NLO EW

Gluon-gluon fusion (ggH). In the ggH production mode, the Higgs boson is produced in an interaction of two gluons via a quark loop (the dominating contributions to the process are from the heavy top and bottom quarks). The final state of a ggH event contains only the decay products of the Higgs boson itself. However, it is possible that one or multiple gluons are radiated from the initial state particles, leading to the presence of additional jets in the final state (the same holds for the other production modes as well).

Vector boson fusion (VBF). In the VBF production mode, the Higgs boson is created in an interaction of two vector bosons (two Z bosons, or W^+ and W^-), emitted by the initial state quarks. In addition to the decay products of the Higgs boson, the final state of a VBF event is characterized by a pair of quarks, which after undergoing hadronization can be observed as a pair of hadron jets. These jets typically have a large invariant mass and angular separation, which is useful for distinguishing VBF events from other events with a similar final state topology.

Higgs-strahlung (VH). In the Higgs-strahlung process, the Higgs boson is emitted by a vector boson (Z or W). The final state in this production mode includes the decay products of the vector boson, which may be a pair of charged leptons (for Z boson decays), a charged lepton and a neutrino (for W boson decays), or a pair of quarks (for both Z and W boson decays).

Production in association with a top quark-antiquark pair ($t\bar{t}H$). In the $t\bar{t}H$ production mode [45, 46], each of the top quarks in the final state instantly decays into a bottom quark and a W boson, and each of the W bosons then produces either a pair of quarks, or a lepton and a neutrino. The $t\bar{t}H$ events therefore occur with a variety of final state topologies, all of which include at least two hadron jets produced by bottom quarks (*b-jets*). Due to their relatively long lifetime of $\sim 10^{-12}$ s, bottom quarks travel a sizeable

³ $\uparrow$ N3LO QCD denotes that the cross-section is calculated at next-to-next-to-next-to-leading order in perturbative QCD describing emissions of gluons by initial state, final state, or intermediate particles.

⁴ $\uparrow$ NLO EW denotes that the cross-section is computed at next-to-leading order in perturbative QED describing photon emissions.

distance (up to several millimeters) before decaying, therefore the b -jets typically feature a *secondary vertex* displaced from the location of the main interaction.

1.3.2 Higgs boson decay modes

With a lifetime of only $\sim 10^{-22}$ seconds, the Higgs boson decays almost instantly after its production. The possible decay modes and their relative rates (*branching ratios*) are defined by the strength of the Higgs field interactions with the SM particles. The channels with the highest branching ratios are shown in Figure 1.4, and the values of branching ratios for $M_H = 125$ GeV are given in Table 1.3. Once again, the fortunate fact that the SM Higgs boson mass is close to 125 GeV makes all of the main decay modes accessible for analysis at the LHC.

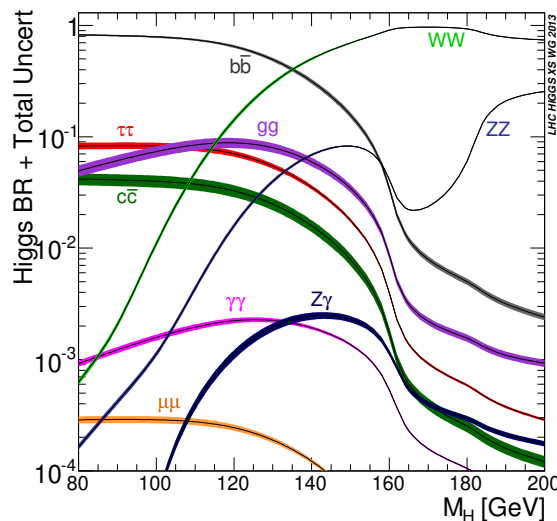


Figure 1.4. Branching ratios of the Higgs boson decay modes [42].

For the fermion-antifermion decay modes, as well as for $H \rightarrow WW$ and $H \rightarrow ZZ$, the cross section of the decay depends only on one Higgs coupling — the coupling to the corresponding SM field. This provides a direct way to establish the value of that coupling experimentally, by measuring the cross section of the respective decay.

Other decay modes, like $H \rightarrow gg$ and $H \rightarrow \gamma\gamma$, include more than one process contributing to their total cross section in the leading order — for instance, the di-photon decay may

Table 1.3. Branching ratios of the Higgs boson decay modes for $M_H = 125$ GeV [42].

Decay channel	Branching ratio	Latest experimental results
$H \rightarrow b\bar{b}$	5.824×10^{-1}	[47, 48]
$H \rightarrow WW$	2.137×10^{-1}	[49]
$H \rightarrow gg$	8.187×10^{-2}	
$H \rightarrow \tau\tau$	6.272×10^{-2}	[50, 51]
$H \rightarrow c\bar{c}$	2.891×10^{-2}	[52, 53]
$H \rightarrow ZZ$	2.619×10^{-2}	[54, 55]
$H \rightarrow \gamma\gamma$	2.270×10^{-3}	[56, 57]
$H \rightarrow Z\gamma$	1.533×10^{-3}	[58, 59]
$H \rightarrow \mu\mu$	2.176×10^{-4}	[1, 60]

occur via a top quark loop or via a W boson loop. In such cases, even the precise knowledge of the cross section is not enough to derive the individual values of the Higgs couplings. However, a combined measurement of multiple decay modes allows to establish the values of the couplings indirectly.

The strength of the Higgs couplings to the SM particles defines the rates at which different decays can be produced in an experiment, but the amount of experimental data is not the only factor affecting the statistical power of the measurement. The efficiency of the detection and identification of the decay products, measuring their parameters, and distinguishing the Higgs decays from other processes with similar topologies depend strongly on the type of particles in the final state. Stable and long-lived particles like photons, electrons, and muons are relatively easy to detect, therefore the Higgs decay modes containing these particles in the final state can be efficiently distinguished from other processes. On the other hand, the hadron jets produced by quarks and gluons are notoriously hard to model and reconstruct, thus, the decay modes containing them in the final state are subject to a less efficient event identification and higher uncertainties. Neutrinos, which appear in many of the final states of the Higgs decays, cannot be directly detected at all.

The statistical significance of the Higgs boson discovery at the LHC in 2012 [61, 62] was dominated by the decay modes with high branching ratios and clean final state signatures: $H \rightarrow ZZ \rightarrow 4\ell$, $H \rightarrow WW \rightarrow 2\ell 2\nu$, $H \rightarrow \gamma\gamma$. The measurements performed in these decay channels allowed to establish the Higgs couplings to W and Z bosons, and to the top quark.

The discoveries of the Higgs boson decays into a pair of bottom quarks [47, 48], and into a pair of τ -leptons [63, 64] have completed the measurements of the Higgs couplings to the third generation fermions.

The first evidence of the $H \rightarrow \mu\mu$ decay with an observed significance of 3.0 standard deviations [1] has marked the beginning of the measurements of the Higgs couplings to second generation fermions. The low branching ratio of the $H \rightarrow \mu\mu$ decay is compensated by the clean signature of the final state, featuring a pair of muons with high transverse momenta, which could be detected and identified with a very high efficiency.

The Higgs boson coupling to c quarks is more challenging to establish due to a limited efficiency of reconstruction and identification of charm jets, despite the $H \rightarrow c\bar{c}$ branching ratio being two orders of magnitude higher than that of the dimuon decay. However, the CMS and ATLAS experiments at the LHC are already achieving a considerable sensitivity in the $H \rightarrow c\bar{c}$ channel [52, 53], hinting at a statistically significant result in the near future. Another decay mode approaching a significant sensitivity at the LHC experiments is the Higgs boson decay into a Z boson and a photon [58, 59].

Establishing the remaining second generation Higgs couplings through a search for the $H \rightarrow s\bar{s}$ decay does not benefit from either a clean final state signature, or from a high branching ratio, therefore it is even more challenging and not realistically possible with currently operating experiments.

The branching ratios of the Higgs boson decays to the first generation fermions are orders of magnitude lower than those for the second generation fermions. Assuming that the Standard Model description of the Higgs couplings is correct, the cross sections of the Higgs boson decays into electrons or first generation quarks are so low that in all years of the LHC operation such events have occurred no more than several times. The studies of such decays may be possible only in the far future with experiments producing Higgs bosons at much higher rates.

1.3.3 Natural width of the Higgs boson

The width of the Higgs boson in the SM is predicted to be approximately 4 MeV [43], while the limited resolution of the CMS and ATLAS detectors allows to experimentally measure the Higgs boson mass with $\sim 1\%$ (~ 1 GeV) precision at best. Therefore, a direct measurement of the natural Higgs width is not possible with the existing detectors. However, the width can be measured indirectly, by exploiting the correlation between on-shell and off-shell⁵ production of the Higgs boson.

The dependency of the on-shell and off-shell production on the strong and weak coupling constants g_s and g , and on the Higgs width Γ_H is as follows:

$$\sigma_{\text{on-shell}} \propto \frac{g_s^2 g^2}{\Gamma_H}, \quad \sigma_{\text{off-shell}} \propto g_s^2 g^2.$$

Therefore, the Higgs natural width can be measured from the ratio of the on-shell and off-shell cross sections. The invariant mass of the off-shell Higgs boson is distributed in the 200–1000 GeV range, where the total background is smaller than that at $M_H = 125$ GeV, which allows to detect the off-shell signal even though it does not form a peak in the mass spectrum.

In a recent analysis [65] by the CMS collaboration, a combined fit of the on-shell and off-shell signals in the $H \rightarrow 4\ell$ channel was performed, and the Higgs width was measured to be $3.2^{+2.4}_{-1.7}$ MeV, in agreement with the SM prediction.

⁵↑ in *off-shell* production, the kinematic relation $E^2 = p^2 + m^2$ is not satisfied for the virtual Higgs boson, therefore the invariant mass of its decay products is not required to be ≈ 125 GeV.

2. EXPERIMENTAL SETUP

As our understanding of the fundamental structure of the world progresses, the experimental equipment needed to probe new hypotheses becomes increasingly more complex and expensive. Since the second half of the 20th century, many of the particle physics experiments have been conducted using *particle colliders*, where high-energy interactions can be produced at high rates. Some of the most ambitious projects in high energy physics have been carried out by the *European Organization for Nuclear Research* (CERN), founded in 1954 by twelve European countries. CERN has hosted numerous experiments, which produced some of the key discoveries in particle physics, such as the discoveries of the W and Z bosons [66–69] at the Super Proton Synchrotron (SPS) in 1983. However, the most notable achievement of CERN was the discovery of the Higgs boson [61, 62] in 2012 at the *Large Hadron Collider* (LHC), which has been CERN’s main accelerator complex since the late 2000’s.

The data used for the studies presented in this thesis has been collected in 2016–2018 by one of the four particle detectors at LHC — the *Compact Muon Solenoid* (CMS). The following sections describe the design, specifications and the performance of both the LHC and the CMS detector.

2.1 The Large Hadron Collider

The Large Hadron Collider [70, 71] is a proton-proton collider situated at the Swiss-French border in the vicinity of Geneva. The original LHC design proposal was approved in 1994; the construction began in 1998 and concluded in 2008. As of 2022, the LHC is the most powerful particle accelerator in the world, and the only collider powerful enough to produce and study processes involving the Higgs boson.

The LHC is a circular accelerator with 27 km in circumference, located in an underground tunnel formerly occupied by the Large Electron-Positron Collider (LEP) [72]. The accelerator ring consists of 1232 superconducting dipole magnets used to curve the trajectories of the proton beams, 474 quadrupole magnets used to focus the beams, and 16 radiofrequency cavities that accelerate the protons. Along the accelerator ring, there are four points in

which the proton beams intersect and collide, and the collisions are recorded by four particle detectors — ATLAS, CMS, ALICE, and LHCb.

Two of the detectors — ATLAS (A Toroidal LHC ApparatuS) [73] and CMS (Compact Muon Solenoid) [74] — are general-purpose detectors designed to probe a wide range of particle physics theories. The most important scientific result achieved with the ATLAS and CMS detectors was the discovery of the Higgs boson [61, 62]. The LHCb experiment (Large Hadron Collider beauty) [75] is focused on b -physics, measuring CP-violating effects in processes involving b -quarks, while ALICE (A Large Ion Collider Experiment) [76] studies the collisions of heavy ions. The schematic depiction of the LHC and its main experiments is shown in Figure 2.1.

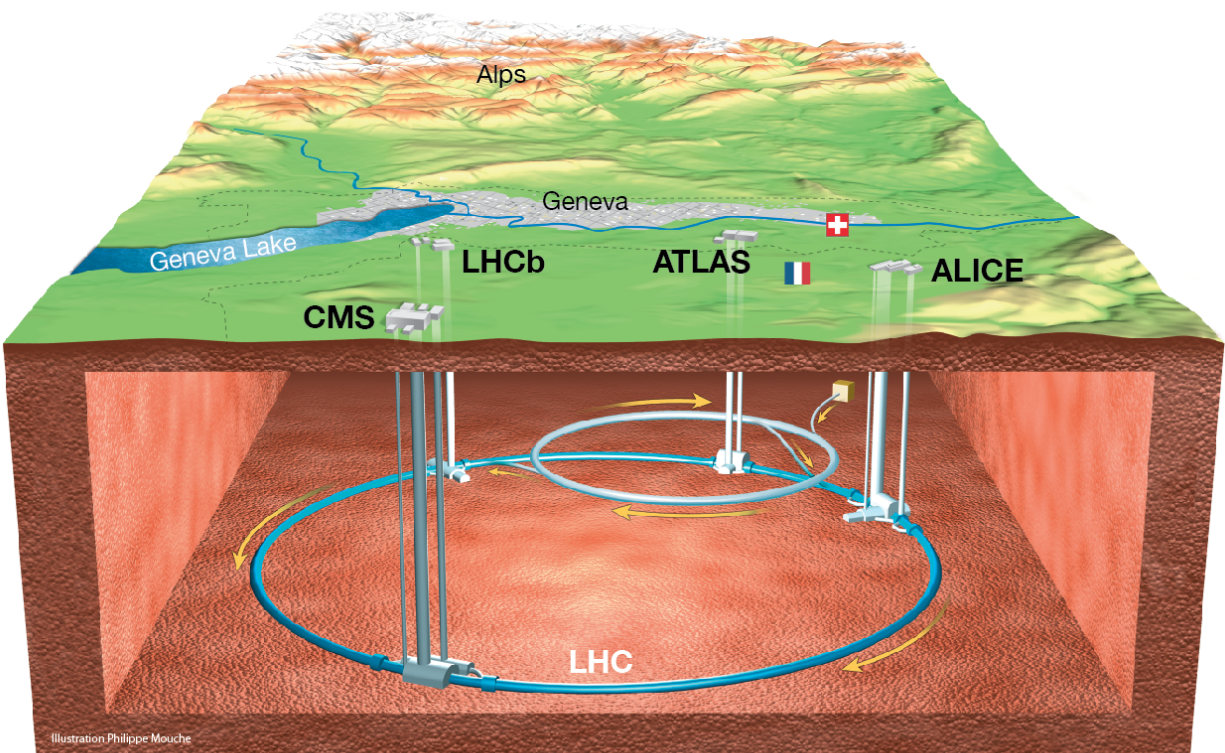


Figure 2.1. Overall view of the LHC [77].

2.1.1 The CERN accelerator complex

The LHC was designed to accelerate either protons or heavy ions (such as Pb, Xe) and collide them at center-of-mass energies up to 14 TeV. In the original design, the protons originated from a container with hydrogen gas, after which they were separated from electrons by an electric field, and then injected into the linear accelerator LINAC2 [78]. Since the commissioning of LINAC4 [79] in 2020, the source of the protons has changed to negative hydrogen ions (H^- , a hydrogen atom with an extra electron). The negative ions are accelerated by LINAC4 to 160 MeV, and the nuclei are separated from the electrons only at injection to the next stage, the Proton Synchrotron Booster (PSB). After that, the protons go through a sequence of accelerators with increasing energy:

- the Proton Synchrotron Booster (PSB): $160 \text{ MeV} \rightarrow 2 \text{ GeV}$,
- the Proton Synchrotron (PS): $2 \text{ GeV} \rightarrow 26 \text{ GeV}$,
- the Super Proton Synchrotron (SPS): $26 \text{ GeV} \rightarrow 450 \text{ GeV}$,

and finally get injected into the LHC, where their energy can be increased up to 7 TeV. The full scheme of the CERN accelerator complex is shown in Figure 2.2 (the scheme depicts the design as of 2013, without LINAC4).

The LHC design allows to simultaneously accelerate up to 2808 proton bunches, with around 115 billion protons in each bunch. The acceleration and grouping into bunches is performed by radiofrequency cavities (RF). The electric field in the RFs oscillates with a given frequency, tuned such that the protons of a particular energy pass through an RF undisturbed, the protons with lower energies are accelerated, and the protons with higher energies are decelerated. The proton bunches are grouped into “trains” (typically, 48 bunches per train), which travel through the LHC ring with temporal separation of 25 ns.

CERN's Accelerator Complex

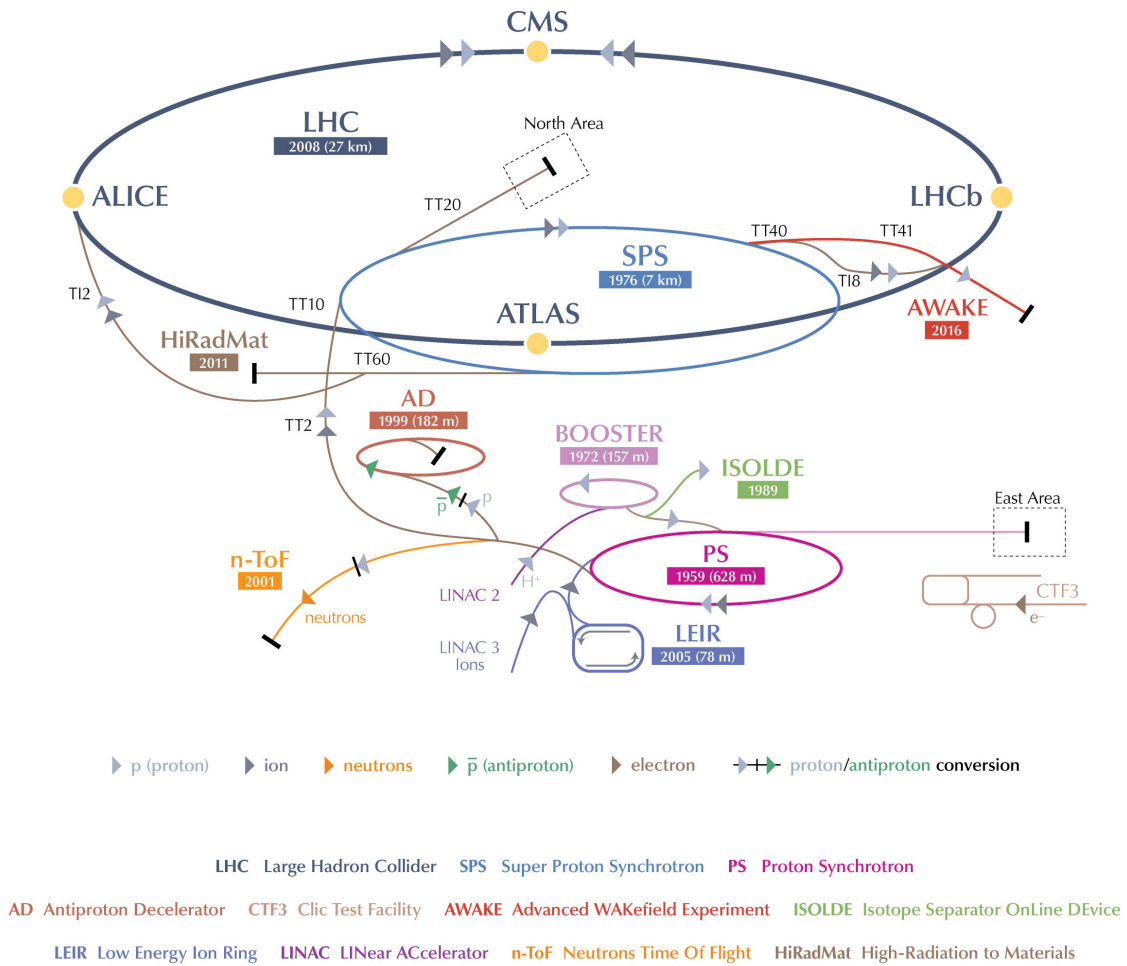


Figure 2.2. Accelerator complex at CERN (as of 2013) [80].

2.1.2 Instantaneous and integrated luminosity

The collision rate of the proton beams at the LHC is characterized by *instantaneous luminosity*, defined as follows:

$$\mathcal{L} = \frac{N_p^2 n_b f \gamma_r}{4\pi \epsilon_n \beta^*} F, \quad (2.1)$$

where N_p is the number of protons per bunch, n_b is the number of bunches, f is the revolution frequency of the bunches, and γ_r is the relativistic factor. Parameters ϵ_n and β^* characterize the focusing of the proton beam, while F is a factor that accounts for the luminosity reduction due to the crossing angle of the beams at the interaction point. The values of these parameters for the nominal LHC design are given in Table 2.1.

Table 2.1. Nominal parameters of the LHC accelerator.

Symbol	Parameter	Nominal value
$\mathcal{L}$	Instantaneous luminosity	$1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$
N_p	Number of protons per bunch	1.15×10^{11}
n_b	Number of bunches per beam	2808
f	Revolution frequency	11245 Hz
ϵ_n	Transverse emittance	3.75 mm μm
β^*	Beta function	0.55 m

The expected event rate $\partial N / \partial t$ for a given process can be obtained from the instantaneous luminosity $\mathcal{L}$ and the process cross section σ using the following relation:

$$\frac{\partial N}{\partial t} = \mathcal{L} \times \sigma. \quad (2.2)$$

Integration over time yields the total number of events produced by a process in a given time period:

$$N = \int \mathcal{L} dt \times \sigma = L \times \sigma. \quad (2.3)$$

The parameter L is called *integrated luminosity*, it characterizes the amount of data produced by the LHC, and it is usually expressed in units of inverse picobarns (pb^{-1}) or femtobarns (fb^{-1}), where $1 \text{ barn} = 10^{-28} \text{ m}^2$. The integrated luminosity recorded by the CMS experiment in each data-taking year from 2010 to 2022 is illustrated in Figure 2.3.

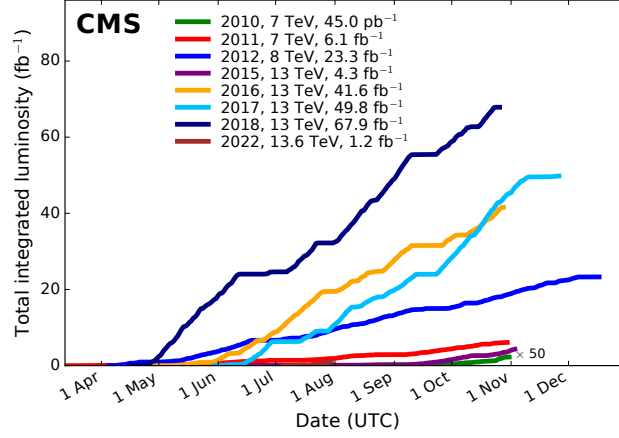


Figure 2.3. Integrated luminosity recorded by the CMS detector during LHC Runs 1 and 2 [81].

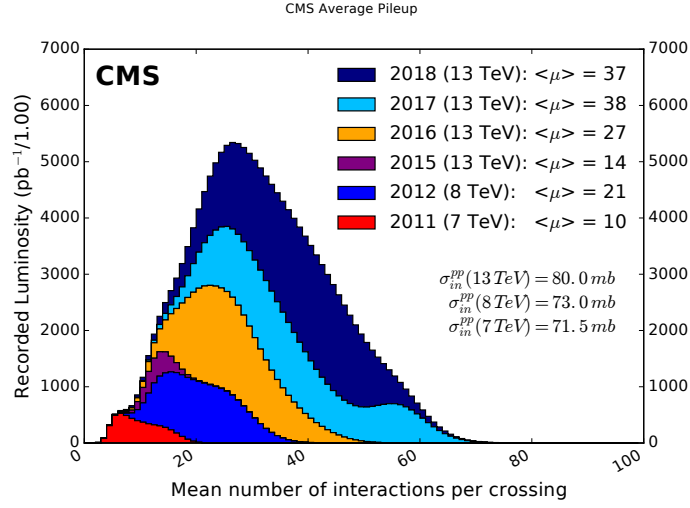


Figure 2.4. Distributions of the number of pile-up interactions in each data-taking period [81].

2.1.3 Pile-up

Due to the large number of protons in the colliding bunches, the collision events in the LHC detectors typically involve more than one interaction at a time. The particles originating from these additional collisions are referred to as *pile-up*. The pile-up is quantified

by the average number of interactions per collision, which can be derived from the LHC parameters as follows:

$$\langle \text{PU} \rangle = \frac{\mathcal{L} \sigma_{pp}^{\text{inel.}}}{n_b f}, \quad (2.4)$$

where $\sigma_{pp}^{\text{inel.}}$ is the inelastic proton-proton cross section, which amounts to approximately 69 mb at $\sqrt{s} = 13$ TeV [82]. High amounts of pile-up present a challenge for the analysis of the collision data, as they increase the occupancy of the detectors, and the response from a pile-up particle can be indistinguishable from that of a particle from a hard collision. The distributions of the number of interactions per collision event for the CMS experiment are shown in Figure 2.4.

2.1.4 The LHC timeline

Since the beginning of its operation in 2010, the LHC has completed two data-taking periods, Run 1 and Run 2, each spanning over several years, and in July 2022 the Run 3 has started. With the start of the LHC Run 3, the center-of-mass energy of the proton beams was increased to the highest value ever, 13.6 TeV. After several years of data collection at this energy (see Figure 2.5), the LHC will be stopped again to undergo an upgrade aimed to increase the instantaneous luminosity by a factor of 5 or more. The upgraded *High Luminosity LHC* (HL-LHC) will operate at $\sqrt{s} = 14$ TeV for the Runs 4 and 5.

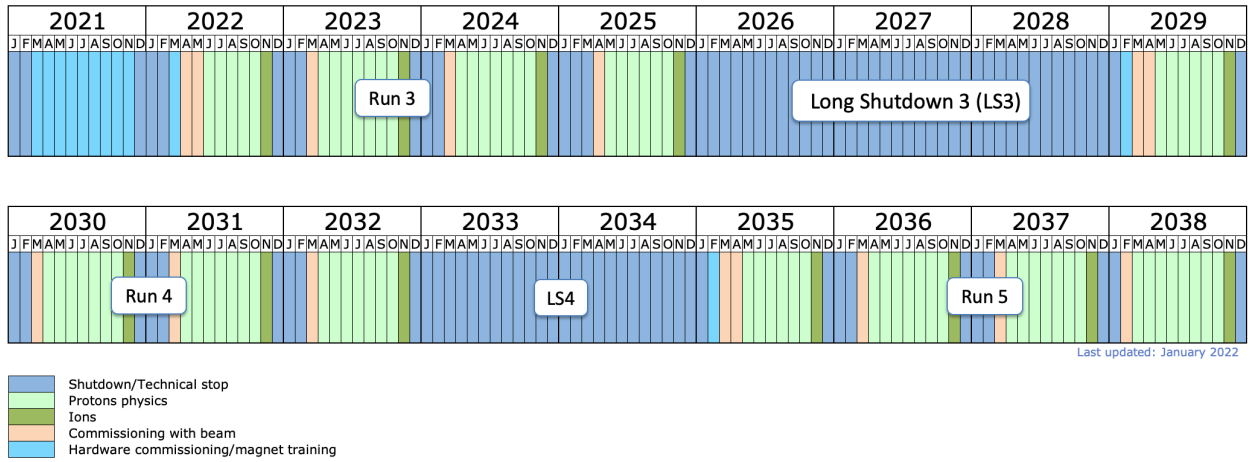


Figure 2.5. LHC timeline for Run 3 and HL-LHC (Runs 4 and 5 [83].

2.2 The CMS Experiment at the LHC

The Compact Muon Solenoid (CMS) detector [74] is one of the four major experiments at the LHC. The multi-purpose CMS detector was designed to probe a wide range of physics theories, from precise measurements of SM parameters, to supersymmetry searches, and dark matter searches.

The schematic view of the CMS detector is shown in Figure 2.6. The central part of the detector is a superconducting solenoid, with various subsystems layered both inside and outside of it. The subsystems cover almost the entire solid angle, with the exception of the regions directly adjacent to the beam pipe.

The layers of the detector, as well as the respective readout electronics, are arranged into three separate parts, which makes it possible to access the inner layers for diagnostics and upgrades without the need to take apart the entire machine. The central cylindrical part is referred to as *barrel*, while the cone-shaped volumes aligned perpendicularly to the beam axis are referred to as *endcaps*.

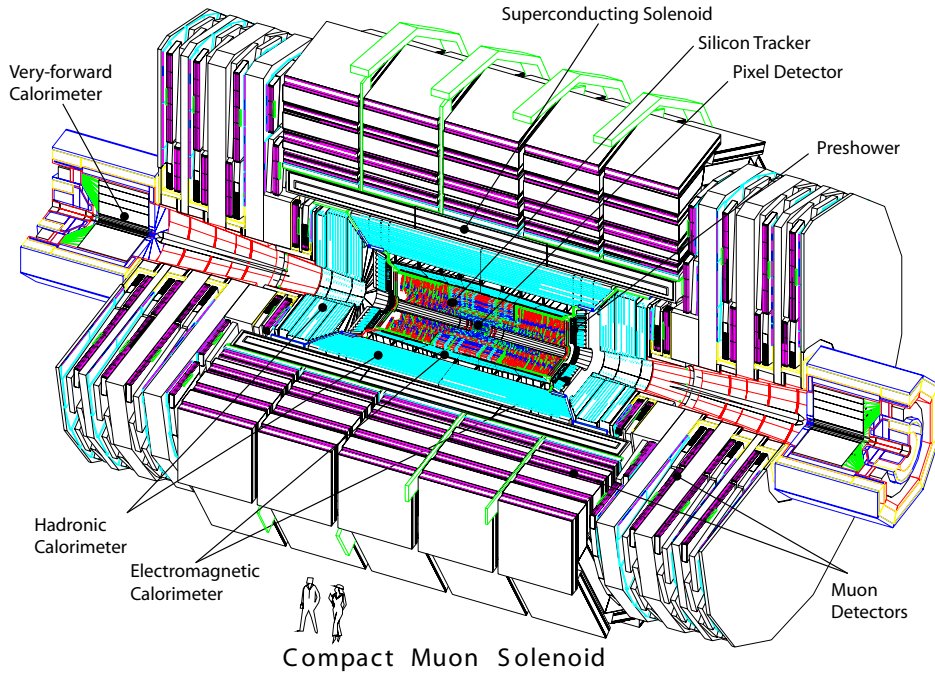


Figure 2.6. A view of the CMS detector and its subsystems [74].

2.2.1 The coordinate system

The coordinate system in the CMS detector is defined based on the LHC ring geometry. The x -axis points towards the center of the LHC ring, the y -axis points upwards, and the z -axis points counter-clockwise along the beam direction. Due to the cylindrical design of the CMS detector, the polar coordinate system is also widely used. In polar coordinates, the azimuthal angle ϕ is measured in $x - y$ plane from the x -axis, the polar angle θ is measured from the z -axis, and the radial distance r is defined as $r = \sqrt{x^2 + y^2}$.

Although both of the colliding proton beams at the LHC have the same nominal energy, the center-of-mass reference frame does not necessarily coincide with the laboratory frame for a given collision. This happens due to the fact that the protons are not fundamental particles, and the primary interaction in fact occurs between *partons* — quarks and/or gluons, which may have different energies. As a consequence, the parton-parton subsystem is often subject to a Lorentz boost along the z -axis.

Instead of cartesian and polar coordinates, it is more convenient to measure quantities that are by definition invariant with respect to Lorentz transformations. Such quantities include the transverse component of the momentum (denoted as p_T), as well as the azimuthal angle ϕ . The third component in this coordinate system is rapidity y , defined as

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

Rapidity is not Lorentz invariant by itself, but it is simply additive; therefore, the Lorentz invariance condition holds for the *rapidity gap* Δy between two particles subjected to the same boost. For ultra-relativistic particles ($E \approx p$), the rapidity can be approximated by *pseudorapidity* η :

$$\eta = -\ln \tan \frac{\theta}{2}.$$

Another useful Lorentz invariant quantity is the *angular distance* between two objects, which is defined as $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. The angular distance is commonly used to define isolation of collision products, and to group collimated particles into *jets*.

2.2.2 The magnet

The name of the CMS detector refers to its central part — a superconducting solenoid made from niobium-titanium alloy and weighing 200 tons. With 12.5 m in length and 6 m in diameter, the magnet is the largest of its kind.

The solenoid generates an almost uniform magnetic field of 3.8 T inside its volume, which is instrumental for determining the charges and momenta of the particles traveling through the detector, since a charged particle in a magnetic field is subjected to the Lorentz force $\vec{F}_L = q(\vec{v} \times \vec{B})$. A stronger field leads to a more curved particle trajectory, which allows to measure the momentum more precisely.

Outside of the magnet, the iron support structure of the detector acts as a return yoke, returning the magnetic flux and reducing the stray magnetic field. The magnetic field in the outer layers of the detector is around 2 T. A detailed map of the magnetic field strength is shown in Figure 2.7.

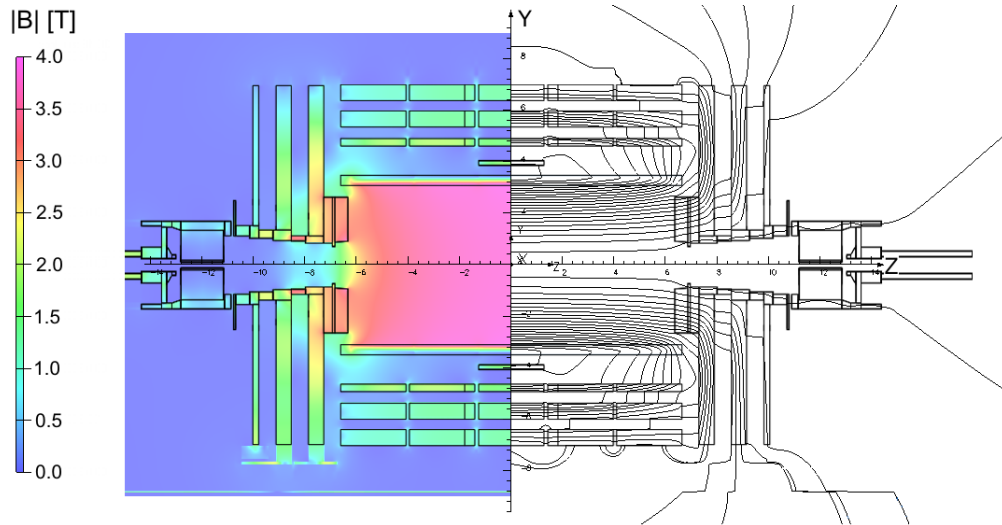


Figure 2.7. The map of the magnetic field strength (left), and the field lines (right) in the CMS detector [84].

2.2.3 The tracker

The innermost subsystem of the CMS detector is the tracker system [85,86]. It is designed to accurately reconstruct the trajectories of charged particles in the close proximity of the collision point, where the detector occupancy is the highest. Another important purpose of the tracker is to reconstruct the secondary vertices displaced from the collision points in the transverse direction; such vertices are characteristic for heavy-flavour processes involving b -quarks or τ -leptons.

The tracker system covers the pseudorapidity range up to $|\eta| = 2.5$ and consists of two main parts: the *pixel tracker*, and the *silicon strip tracker*. The layout of the tracker in the original design is shown in Figure 2.8. Both the pixel and the strip detectors use silicon semiconductors for particle detection, due to their compact size and fast readout. In such detectors, a charged particle passing through the sensor creates an electron-hole pair, which then generates a readout signal as the electrons drifts towards the electrodes. The main difference between the pixel and the strip trackers is in their geometry: a single pixel module provides three-dimensional spatial information, whereas a single silicon strip provides only two-dimensional information. However, two silicon strips overlapping at an angle can provide three-dimensional information as well.

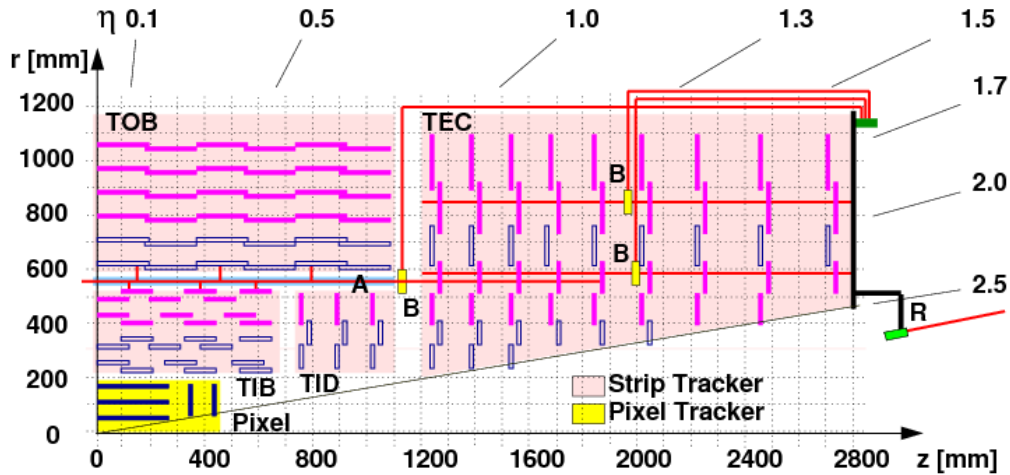


Figure 2.8. Original layout of the CMS tracker [87].

The pixel detector is located closest to the collision point, where a higher granularity is necessary. In the original design, the pixel detector featured three barrel layers and two endcap disks on each side, with a total number of readout channels around 66 million. After the Phase-1 upgrade in March 2017 [88], the number of readout channels was increased to 124 million, and the number of layers was increased as well (see Figure 2.9). In the upgraded detector, the pixels are arranged into four cylindrical barrel layers, with radii ranging from 2.9 to 16 cm, and three endcap disks on each side, which cover radii from 4.5 to 16.1 cm. The endcap disks are located at a distance of 21.9–51.6 cm from the collision point in the longitudinal direction. Due to the high granularity (a single sensor is $100 \times 150 \mu\text{m}^2$ in size), the pixel detector features an excellent spatial resolution of $10 \mu\text{m}$ in the transverse direction, and $15 \mu\text{m}$ in the longitudinal direction, while the overall efficiency of the hit reconstruction is around 99.95% [89].

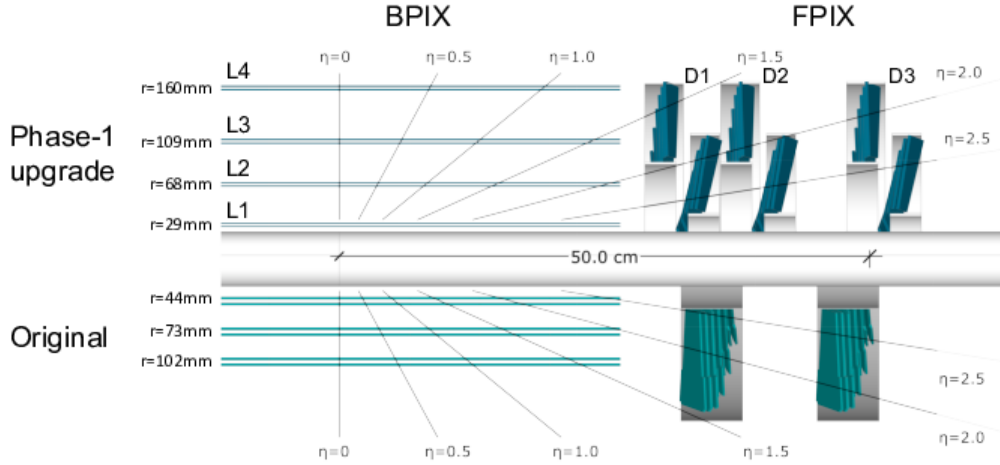


Figure 2.9. Geometry of the pixel detector before and after the Phase-1 upgrade [88].

The silicon strip tracker is positioned outside of the pixel detector, and consists of two parts: the inner strip tracker, and the outer strip tracker. The inner strip tracker consists of *tracker inner barrel* (TIB), which features four layers covering radii from 20 to 55 cm from the beam, as well as three layers of *tracker inner disks* (TID) on each side. The outer strip tracker features the *tracker outer barrel* (TOB), containing six layers with radii up to 1.1 m, and the *tracker endcaps* (TEC), containing nine disks on each side. Both inner and outer

parts of the strip detector contain single-sided and double-sided strips. In the double-sided modules, the strips are rotated by $100\ \mu\text{rad}$ with respect to each other, which allows to obtain a three-dimensional position of a hit.

The spatial resolution of the strip tracker in transverse direction ranges from 23 to $34\ \mu\text{m}$ in TIB and TID, and from 35 to $52\ \mu\text{m}$ in TOB and TEC. The resolution in the longitudinal direction is approximately 10 times larger than that, while the overall hit efficiency of the strip tracker is 99.8%.

2.2.4 The electromagnetic calorimeter

The next detector layer outside of the tracker is the electromagnetic calorimeter (ECAL) [90], the purpose of which is to measure the energy of electromagnetic showers induced by electrons and photons. As the ECAL is the only CMS subsystem suitable for photon detection, a high granularity and a good energy resolution are required.

The ECAL design features homogeneous scintillator modules made from lead tungsten (PbWO_4) crystals. The lead tungsten acts as both absorber, inducing the electromagnetic shower, and active material, producing the scintillation light. The scintillation signal is collected and amplified by silicon avalanche photodiodes in the barrel and by vacuum phototriodes in the endcap regions. The benefits of PbWO_4 are its high density ($\rho = 8.28\ \text{g/cm}^3$), low radiation length ($X_0 = 0.89\ \text{cm}$), and fast light yield (signal readout takes 5–10 ns). Each crystal covers a solid angle of approximately 0.0174×0.0174 radians, and the entire ECAL, with 61200 crystals in the barrel and 7324 crystals in the endcaps, covers the pseudorapidity range up to $|\eta| = 3$. The depth of the crystals equals around $25X_0$, which ensures that the electromagnetic showers with energies up to 1 TeV are on average 98% contained for both electrons and photons. The axes of the crystals are tilted by up to 3° with respect to the direction to beamspot, in order to avoid loss of data due to possible gaps between the modules. In the endcap region, the ECAL is augmented by a lead/silicon-strip preshower (ES) detector, covering the pseudorapidity range $1.6 < \eta < 2.6$. The schematic view of the ECAL and ES is shown in Figure 2.10.

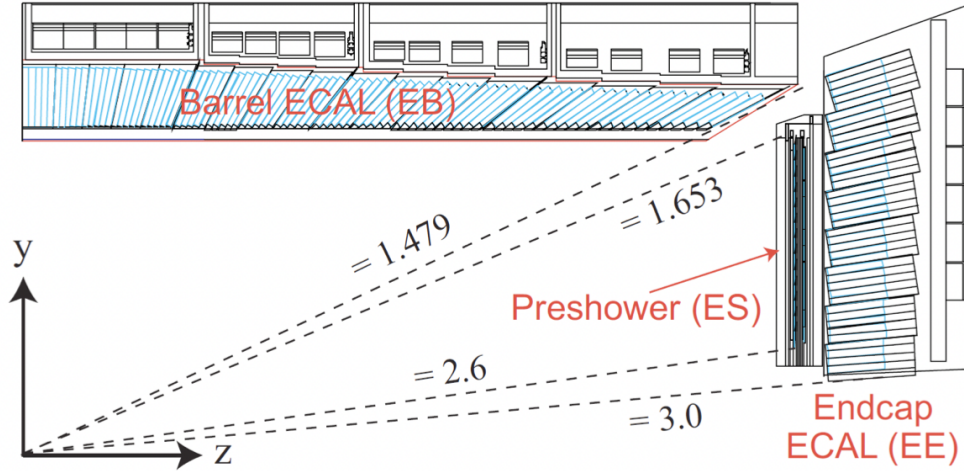


Figure 2.10. Schematic view of the CMS ECAL [91].

The energy resolution of the ECAL can be parameterized as

$$\frac{\sigma_E}{E} = \frac{a}{E} \oplus \frac{b}{\sqrt{E}} \oplus c,$$

where a characterizes electronics noise and pile-up, b is a stochastic term characterizing fluctuations in event-by-event measurements, and c is a constant term related to the energy scale calibration. For PbWO_4 crystals, the measured energy resolution is

$$\frac{\sigma_E}{E} = \frac{12\%}{E/\text{GeV}} \oplus \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus 0.3\%.$$

2.2.5 The hadron calorimeter

The last detector layer inside the solenoid is the hadron calorimeter (HCAL) [92]. The HCAL is used to measure the energies of hadron showers induced by both charged and neutral hadrons. Unlike ECAL, the HCAL does not use a homogeneous scintillator design, as the volume that is required to contain a typical hadron shower would be too large to fit inside the magnet. Instead, the HCAL features layered modules, where brass absorber plates alternate with plastic scintillators.

The main part of the HCAL consists of the barrel (HB) and endcap (HE) hadron calorimeters, which cover the pseudorapidity range up to $|\eta| = 3$, and provide a material budget of around 7 radiation lengths in HB and 11 radiation lengths in HE. The scintillator layers are arranged into *towers*, providing coarse granularity of 0.087×0.087 radians in HB, and 0.17×0.17 radians in HE. The readouts of scintillators from the same $\eta - \phi$ tower are combined together, recording only the total energy response without depth granularity.

To improve the hermeticity of HCAL, it is augmented by two more subsystems: the forward calorimeter (HF) and the outer calorimeter (HO). The HF is located 11.2 m away from the collisions point along the z -axis, and covers a pseudorapidity range up to $|\eta| = 5.2$. The HO is positioned outside of the magnet, using the magnet material as an absorber. This additional layer increases the HCAL material budget in the barrel up to 11 radiation lengths, making it comparable to that of HE.

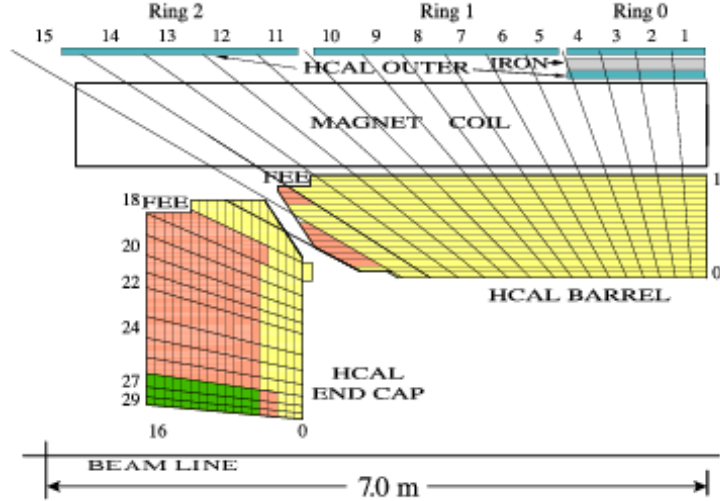


Figure 2.11. Schematic view of the CMS HCAL [93].

The combined energy resolution of ECAL and HCAL was measured to be [94]

$$\frac{\sigma_E}{E} = \frac{84.7\%}{\sqrt{E/\text{GeV}}} \oplus 7.4\%,$$

which is significantly worse than the energy resolution of the ECAL alone. Despite the modest spatial and energy resolution, the wide pseudorapidity coverage of the HCAL allows

to infer the transverse energy carried by undetected particles such as neutrinos, making it instrumental in many physics analyses.

During the Phase-2 upgrade of the CMS detector after Run 3, the ECAL and HCAL modules in the endcaps will be replaced by the High Granularity Calorimeter (HGCAL) [95], which will improve the spatial and energy resolution, as well as add depth granularity.

2.2.6 The muon system

As the name of the CMS detector suggests, efficient detection of muons is one of the experiment's core specifications. Muons produced in the proton-proton collisions have momenta ranging from a few GeV to several hundreds GeV. For such muons, the energy loss as they travel through the detector layers is much lower, in comparison to other particles. This property makes detection and identification of muons quite easy, as no other particles are capable of penetrating the material of the calorimeters and the magnet without losing all energy due to strong or electromagnetic interactions.

The detection of muons in CMS is performed by the muon system [96] layered outside of the solenoid. Since the muon system must cover a large surface area, the choice of the type of sensors was made towards gas ionisation chambers, which provide a decent performance at a reasonable price. There are three types of gaseous detectors featured in the muon system: *drift tubes* (DT), *cathode strip chambers* (CSC), and *resistive plate chambers* (RPC). As shown in Figure 2.12, the DTs are used in the barrel region, the CSCs are used in the endcaps, and the RPC modules are present in both barrel and endcaps.

The DT chambers feature $1.3 \times 4.2 \text{ cm}^2$ cells filled with a mixture of Ar (85%) and CO₂ (15%). A muon traveling through the gas induces ionisation, and the electrodes located at the top and the bottom of a DT cell provide an electric field through which the knocked-off electrons move at a constant drift velocity of $55 \text{ } \mu\text{m/s}$ towards an anode wire. The muon position and angle are then be derived from the drift time, with spatial resolution ranging from 80 to $120 \text{ } \mu\text{m}$.

The CSC chambers feature trapezoidal modules with alternating layers of anode wires and cathode strips, oriented orthogonally to each other. The CSC modules are filled with a

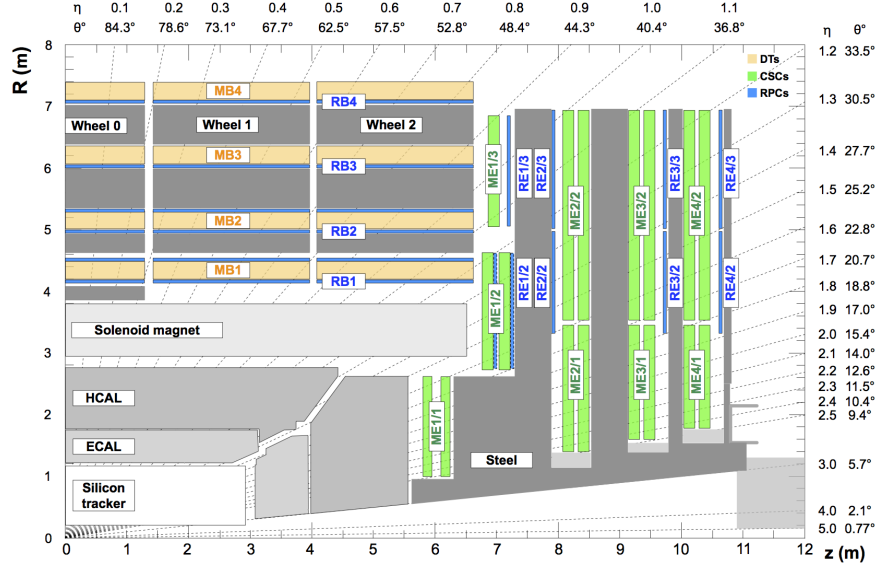


Figure 2.12. Schematic view of the CMS muon system [97].

mixture of Ar (40%), CO₂ (50%), and CF₄ (10%), which gets ionized upon a muon traverse, and the combined signal from strips and wires provides a precise muon position with spatial resolution of 40 – 150 μm . The CSC modules cover the pseudorapidity range up to $|\eta| = 2.4$.

The RPC modules are placed in both barrel and endcap regions. They feature pairs of resistive plastic plates, with the gap between them filled with a mixture of C₂H₂F₄ (95.2%), Iso-C₄H₁₀ (4.5%) and SF₆ (0.3%). The RPCs operate in avalanche mode, with signal collected by metallic strips located on the outer surface of the plates. Unlike DTs and CSCs, the RPCs provide only a coarse (~ 1 cm) spatial resolution. However, their advantage is in excellent time resolution, as the timing precision is better than 3 ns. Thus, RPCs are primarily used for fast muon triggering, whereas the muon trajectory reconstruction relies on the DT and CSC signals.

Overall, the CMS muon system allows to reconstruct and identify muons with $> 96\%$ efficiency and ≈ 1.4 ns timing resolution. The combined momentum resolution of the muon system and the tracker is approximately 3% in the endcaps and 1% in the barrel for muons with momenta up to 100 GeV [97]; the resolution is dominated by the performance of the inner tracker.

2.2.7 The trigger system

At the design luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, the LHC collides 2544 bunches with temporal separation of 25 ns, which translates to a collision rate of 40 MHz. The size of the total raw detector readout for one collision event is of the order of 1 MB, therefore, to accept all collisions, the CMS would need to have the ability to record data at 40 TB/s rate, which is far beyond the capabilities of the existing technologies.

On the other hand, the “interesting” event signatures do not occur in the majority of the proton-proton collisions. The total cross section of a proton-proton interaction at $\sqrt{s} = 13 \text{ TeV}$ is $\sim 10^{11} \text{ pb}$, while the cross sections of the most common SM interactions — production of W and Z bosons — are $\sim 10^5 \text{ pb}$ [98], and the Higgs boson production cross section is $\sim 44 \text{ pb}$ [43].

The CMS trigger system is designed to quickly select the “interesting” events out of the ongoing collisions, and to bring the rate of data recording down by several orders of magnitude. The system consists of the *level-1* (L1) trigger and the *high level trigger* (HLT).

The L1 trigger [99] operates on raw detector readouts, and has a latency of $3.8 \mu\text{s}$. Within this time interval, the trigger performs a rough reconstruction of particles using the information from the calorimeters and the muon system. The signals are combined into *trigger primitives*, which are used to build trigger candidates, such as muons, photons, electrons, jets, and τ -leptons. The total event information is also used to compute the missing transverse energy. After the trigger candidates are built, a set of selections is applied to accept or reject an event based on its topology and the kinematics of the L1 primitives. In order to keep the total L1 rate under control, the L1 trigger paths corresponding to different event topologies are *prescaled*, which means a certain share of events is discarded even if they pass the L1 kinematic selections. In Run 2, the CMS L1 trigger was configured to reduce the event rate from 40 MHz to under 100 kHz.

The high level trigger [100] is used to further reduce the rate to $\sim 1 \text{ kHz}$, at which point the CMS data acquisition system is capable of writing the events to permanent storage. Unlike the L1 triggers, the HLT is software-based, it uses a computing farm with 32 000 CPU cores to process the events accepted by the L1 trigger, reconstruct the particles more

precisely, and apply an additional set of selection. Due to the reduction of the event rate by the L1 trigger, the HLT has more time to process each event — up to $320\ \mu\text{s}$. This allows to perform the reconstruction of the event objects using the information not only from the calorimeters and the muon system, but also from the tracker. After applying the improved reconstruction sequences, the HLT takes a decision either to reject the event, or to save it into one or several *HLT paths*, which correspond to different topologies and kinematics. In Run 2, the CMS HLT menu contained around 600 trigger paths to save the “interesting” events for various physics analyses.

The muon triggers are of a particular interest: the CMS detector has an exceptional muon detection capability, as well as an excellent muon p_T resolution, thus efficient triggering algorithms are essential to use the detector to its full potential. Additionally, the muon selection is an integral part of the $H \rightarrow \mu\mu$ analysis, which is the focus of this thesis. The next chapter describes in detail the design of the CMS muon HLT and the developments that have been made to its algorithms in view of the LHC Run 3.

2.2.8 Event reconstruction

The particles detected by the CMS experiment are identified by the signals they produce in different detector layers. For instance, both a photon and an electron deposit all of their energy in the ECAL, but the electron also leaves a signal in the silicon tracker, while the photon does not. Hadrons and hadron jets are identified by the signals in the silicon tracker and the HCAL, while muons are detected by the muon system in addition to the tracker signals.

The trajectories of charged particles are reconstructed from individual detector signals by grouping them based on spatial and kinematic compatibility. The curvature of the trajectory, along with the knowledge of the magnetic field distribution in the detector, allows to infer the charge of the particle and its momentum¹. The changes in the trajectory curvature after crossing the detector layers also allow to determine the direction in which the particle travels,

¹↑The trajectory of a charged particle in a uniform magnetic field is a helix; in realistic detector conditions, the irregularities of the magnetic field must be taken into account, as well as the energy losses when the particle crosses detector layers.

therefore the particles from the proton-proton collisions can be distinguished from cosmic ray particles.

Once the trajectories of all charged particles are built, they are grouped based on spatial vicinity and form *vertices* — points, in which multiple trajectories intersect. A collision event typically has no more than one vertex corresponding to a “hard” collision in which particles with high transverse momenta are created. At the same time, there may be up to several dozen additional vertices — pile-up — which complicate the event reconstruction.

In the $H \rightarrow \mu\mu$ search, naturally, the most important reconstructed particles are muons. Additionally, since some of the Higgs production modes feature quarks in the final state (see Section 1.3.1), the $H \rightarrow \mu\mu$ search strategy relies on the reconstruction of hadron jets produced by them.

The muon trajectories are reconstructed by combining signals from the silicon tracker and the muon system using the *Kalman filter* algorithm described in Section 3.1. The kinematic properties of the muon are then established by fitting the trajectory according to the theoretical knowledge of the muon propagation in the magnetic field.

The values of the fitted parameters are imprecise due to the limited spatial resolution of the muon system, as well as an imperfect knowledge of the magnetic field distribution and the alignment of the subdetectors. The muon transverse momentum resolution is shown in Figure 2.13. The limited momentum resolution is particularly relevant for the $H \rightarrow \mu\mu$ searches, as it translates into the limited dimuon mass resolution. As a result, the Higgs peak in the dimuon mass spectrum with a natural width of ≈ 4 MeV appears several GeV wide in the detector.

The hadron jets in the CMS detector are built by clustering the reconstructed tracks located at a close angular distance from each other. In the $H \rightarrow \mu\mu$ analysis, as well as in numerous other searches, the anti- k_T algorithm [101] with a distance parameter of $R = 0.4$ is used.

In addition to the jets produced by quarks in the final states of the Higgs production processes, there are typically jets from other sources. In particular, additional jets may be generated by gluons radiated by the hardons participating in the hard collisions. A considerable amount of jets is also produced by the pile-up interactions, or misreconstructed

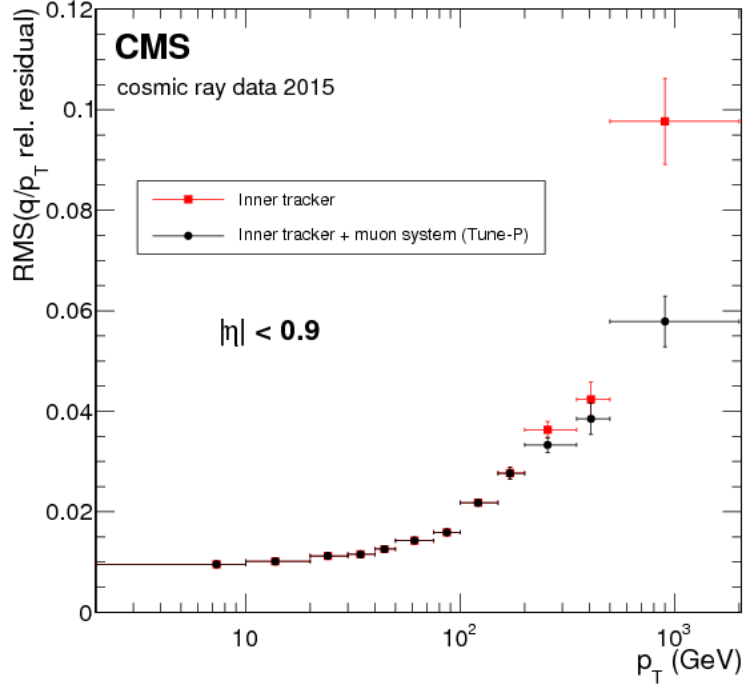


Figure 2.13. The transverse momentum resolution of muons: RMS of q/p_T residuals distribution as a function of p_T measured using cosmic rays [97].

from the detector noise. The pile-up jets typically have lower transverse momenta, and populate the forward region of the detector; however, a presence of a high- p_T pile-up jets is also possible in all pseudorapidity regions.

The reconstruction of jets and determining their kinematic properties is challenging due to the presence of neutral hadrons which cannot be tracked to a particular vertex, and the small angular distance between the tracks of charged particles, which complicates their reconstruction. Because of this, the reconstruction of jets in the detector is usually associated with high uncertainties on their kinematics, which contribute to the systematic uncertainties in the analyses where the jets are used. This is especially important for analyses in which the jets are expected to be in the forward pseudorapidity region, as the contribution of the pile-up there is the highest.

3. CMS MUON HIGH LEVEL TRIGGER

The properties of the muon, such as its relatively high mass and long lifetime, make it the only charged particle capable of passing through large volumes of dense material with minimal energy loss. This allows to detect and identify muons at particle colliders with an exceptional efficiency, by placing a sufficient amount of material between the collision point and the muon detectors.

At the same time, muons are integral for many of the SM interactions — primarily, through the $Z \rightarrow \mu^+\mu^-$ and $W \rightarrow \mu\nu_\mu$ decays, which, in turn, occur in the processes that involve heavy quarks or τ -leptons. The Higgs boson decay into a muon pair ($H \rightarrow \mu\mu$) provides a direct way to measure the Higgs coupling to muons, which is required for the validation of the Standard Model. Various theories attempting to describe the physics beyond the Standard Model (BSM) also frequently include processes with muons in the final state.

As mentioned in Section 2.2.7, the rate of the proton-proton collisions at the LHC exceeds the rate at which the collision event data can be recorded by the CMS experiment by several orders of magnitude. By quickly identifying the collisions with muons in the final state, it is possible to ensure that the “interesting” events, which feature the SM and BSM processes mentioned above, are not discarded during the unavoidable rate reduction.

The CMS muon trigger uses the signals from both the muon system and the silicon tracker to reconstruct and identify the muons. The signals induced by a muon in the detector layers are combined into a trajectory using pattern recognition algorithms, and then fitted with a helical function to extract the kinematic properties. After the muons as well as other objects used for triggering are reconstructed, the event topology is identified and a decision is made whether to keep or discard the event. This general approach is applied twice, first using a hardware *level-1 trigger* (L1), and then using a software *high level trigger* (HLT).

This chapter provides an in-depth discussion of the muon HLT algorithms used in Run 2, as well as a description of a novel approach developed for Run 3.

3.1 Muon tracking, reconstruction and triggering algorithms

All tracking and reconstruction algorithms used in CMS [102] are formulated in terms of the basic concepts, the definitions for which are given below.

Hit.

A charged particle traveling through the detector induces an analogue charge signal each time it crosses a layer of the silicon tracker or one of the muon detectors in the CMS muon system. If the charge exceeds a given threshold, the precise location of the corresponding signal is saved into an object referred to as *hit*. A hit contains either 2D or 3D position information, depending on the geometry of the particular layer. One dimension is associated with the distance from the layer at which the hit was registered to the center of the detector. The additional dimensions are associated with the position of the hit on the layer, which is known either in only one dimension (for example, for the silicon strips in the tracker), or in two dimensions (for example, for the pixel detector, or stereo layers of the silicon strip detectors).

The signals from the muon system are perfect for muon identification, as it is unlikely for any other charged particle to penetrate through the material of the calorimeters and the CMS magnet. In turn, the unparalleled precision of the tracker is crucial for the accurate estimation of the 4-momentum of a muon. However, the tracker is crowded with hits produced by other charged particles, which are indistinguishable from those generated by the muons of interest. The large amount of hits in the silicon tracker, often caused by pile-up, constitutes one of the main challenges in muon reconstruction.

Trajectory.

The *trajectory* of a charged particle is described by a parametrization of the path it took through the CMS detector as well as an associated collection of hits that are assumed to be produced by this particle. A trajectory can combine hits from different subdetectors.

The trajectory of a charged particle in a magnetic field is determined by Maxwell's equations, in particular the following differential equation:

$$\frac{d^2\vec{r}}{ds^2} = \frac{q}{p} \frac{d\vec{r}}{ds} B(r), \quad (3.1)$$

where $\vec{r}(s)$ is the position of the particle along its trajectory, which is parameterized with the arc length parameter s , $\frac{d^2\vec{r}}{ds^2}$ is the curvature of the trajectory, and $\frac{d\vec{r}}{ds}$ is the unit tangent to the trajectory. q/p is the particle charge divided by the magnitude of its momentum, and $B(r)$ is the magnetic field. The solution to this equation is a helix parameterized with five variables. The particular choice of the parameters depends on the purpose and the point of the trajectory where the parameters are defined [103]. Two approaches to the trajectory parameterization are usually considered: using *global parameters*, defined at the point of the closest approach of the trajectory to the beamline, and *local parameters*, defined at the point where the trajectory crosses a detector layer.

The global parameters are d_0 , z_0 , $\cot\theta$, ϕ and p_T . The first two, d_0 and z_0 , are the transverse and longitudinal distances from the point of the closest approach of the trajectory to the assumed position of the particle production point. The parameters ϕ and θ are the azimuthal and polar angles of the momentum vector, and p_T is the magnitude of the transverse component of the momentum.

The local parameters (q/p , λ , ϕ , x_T and y_T) are defined in the curvilinear frame. Here λ is the dip angle of the particle's 3-momentum, x_T and y_T are defined in the plane perpendicular to the trajectory of the particle at a given point, such that x_T also lies in the global xy -plane, and ϕ is the azimuthal angle in the global xy -plane. The local parameterization at a given detector layer is referred to as a *trajectory state on surface* (TSOS).

Uncertainties on the trajectory parameters (in any parameterization) and the correlations between them are provided as a 5×5 covariance matrix associated with the trajectory.

Trajectory building.

Trajectories are built using a Combinatorial Track Finder (CTF) algorithm [102] based on the *Kalman filter* method [104]. The process includes the following steps:

1. **Seed generation.** The initial estimate of the trajectory parameters is provided as a *seed* — an object that contains an estimate of the trajectory parameters and their uncertainties, as well as any hits used in the creation of the seed.
2. **Trajectory candidate building.** A procedure that starts from a seed and iterates over the detector layers compatible with the assumed trajectory. At each iteration, the layer is scanned for compatible hits, with the search window defined by the uncertainties on the trajectory parameters. If there are compatible hits, each of them is used to create a trajectory candidate by updating the original trajectory with that hit, and then proceeding to the next iteration. If no compatible hits are found, the building of the candidate is aborted. However, the algorithm may be configured to allow skipping one or several layers if no compatible hits were found. The maximum number of compatible hits to try is also a configurable parameter in the algorithm.

Once all trajectory candidates in the event have been built, they are compared with each other to determine whether any trajectories share the same hits. The trajectories are then cleaned to ensure that each hit belongs to only one trajectory.
3. **Track producer.** Once a trajectory is fully reconstructed and cleaned, the final fit is performed in an iterative way using the Kalman filter approach again, and the resulting parameterization of the trajectory is referred to as a *track*.
4. **Track filtering.** The tracks are required to pass a set of quality selections, such as a minimum number of hits in the respective trajectory, or a limit on the normalized χ^2 of the fit.

Stages of muon reconstruction.

In CMS, both hardware and software triggers are used to reconstruct muons. The hardware trigger (**L1 trigger**) uses only the signals from muon system (DTs, CSCs, RPCs, as well as GEMs since Run 3). The software trigger (high level trigger, HLT) contains two stages:

- **L2 reconstruction:** building trajectories seeded by filtered L1 muons and using only information from the muon system. A Kalman filter approach similar to the track reconstruction in the silicon tracker is used. While this reconstruction uses the same detector information as the L1 trigger, this approach results in a more precise estimate of the properties of the muon candidates. For prompt muons, these estimates can be further improved by using the beam spot position as a constraint.
- **L3 reconstruction:** combines information from muon system and tracker. The trajectory building in the tracker is seeded by either L2 or L1 muons, and the reconstruction is performed using the Kalman filter approach described above. If the seeding is based on L2 muons, the hits from the tracker track and the matching L2 muon are used to perform a global fit of the track parameters, analogous to “global muons” in the offline reconstruction. If the seeding is based on L1 muons, the tracker tracks are matched to muon segments and if at least one matching segment is found, the track is tagged as a “tracker muon”, again analogous to the approach in the offline reconstruction. The reconstruction can be performed in two ways, depending on the order in which the tracker layers are considered:
 - **Outside-in (OI) reconstruction:** a L2 muon is reconstructed in the muon system, its trajectory is propagated to the outer tracker, and the L3 trajectory is built using the Kalman filter algorithm, which iterates over the tracker layers going from the outer layers inwards.
 - **Inside-out (IO) reconstruction:** initial estimate of the trajectory is given by either L1 or L2 muon reconstructed in the muon system, but the trajectory building in the tracker starts from the innermost layer.

Efficiency definitions for L3 studies.

The reconstruction of muons at each level is not guaranteed to always be successful. There are multiple factors that can lead to a failed reconstruction:

- A muon did not produce a hit on a detector layer.
- A muon produced a hit, but it was not found by the trajectory builder due to inaccurate parameters of the track seed.
- A muon produced a hit, but pile-up hits were more geometrically compatible with the trajectory.

The combined effect of these factors affects the performance of any reconstruction algorithm, which is quantified by *efficiency* of the algorithm — the fraction of successfully reconstructed muons. Two different efficiency measurement techniques were used in the measurements presented in the following:

1. **ΔR matching.** In this approach, the efficiency of the outside-in reconstruction of the muon trajectory in the tracker, “L3(OI) efficiency”, is measured with respect to L2 muons matched to generated muons. The approach is summarized below:
 - All L2 muons used in the efficiency calculation are required to have at least one generated muon within $\Delta R < 0.3$. This ensures that fake L2 muons are excluded from the efficiency calculation.
 - In the requirement above, only generated muons with $|\eta| < 2.4$ are considered.
 - A L2 muon *passes* the L3 reconstruction if there is at least one L3 muon track within $\Delta R < 0.3$ from it.
 - The efficiency is calculated as the number of L2 muons passing L3 reconstruction divided by the total number of L2 muons.

The benefits of the ΔR matching approach are its speed and simplicity of implementation, which makes this efficiency definition convenient to use in various automated optimization studies. This approach cannot be used to compute the L3 efficiency on real data, so it is only used for MC.

2. **Tag-and-probe technique.** To evaluate the efficiency in real data we use a “tag-and-probe” technique [105], which works under the assumption that the event contains a pair of muons from a Z boson decay.

In this approach, one offline muon (“tag”) is required to pass tight selections, and to have fired the isolated single-muon trigger, while another muon from the same event (“probe”) is used to measure the L3 reconstruction efficiency. The invariant mass of the two muons is restricted to a window around the mass of the Z boson.

The implementation of the method closely follows that used in Ref. [106]. The selections for “tag” and “probe” muons are summarized below:

- Tag requirement:
 - Offline reconstructed muon
 - $p_T > 26$ GeV
 - $|\eta| < 2.4$
 - Tight ID and isolation
 - Matched to Iso(Tk)Mu24
- Probe requirement:
 - Offline reconstructed muon
 - $p_T > 24$ GeV
 - $|\eta| < 2.4$
 - Tight ID
 - Tight PF isolation
 - $81 \text{ GeV} < m(\text{tag, probe}) < 101 \text{ GeV}$
 - $\Delta R(\text{tag, probe}) > 0.3$
 - Matched to at least one L1 muon with $p_T > 22$ GeV and quality ≥ 12 with $\Delta R < 0.3$
- Passing probe requirement:
 - Probe matched to an L3 muon track within $\Delta R < 0.3$

It is important to note that the ΔR -based method computes the efficiency with respect to L2 muons, which is needed to correctly estimate the algorithmic efficiency of the reconstruction of L3 muons seeded by L2 muons. In turn, the tag-and-probe method computes the L3 efficiency with respect to L1 muons, which is convenient for calculation of the efficiency of the full HLT sequence, as it contains reconstruction steps seeded by L2 and L1 muons. In practice, the efficiency values obtained with the two methods are close to each other due to the high efficiency of L2 reconstruction, as shown in Figure 3.1.

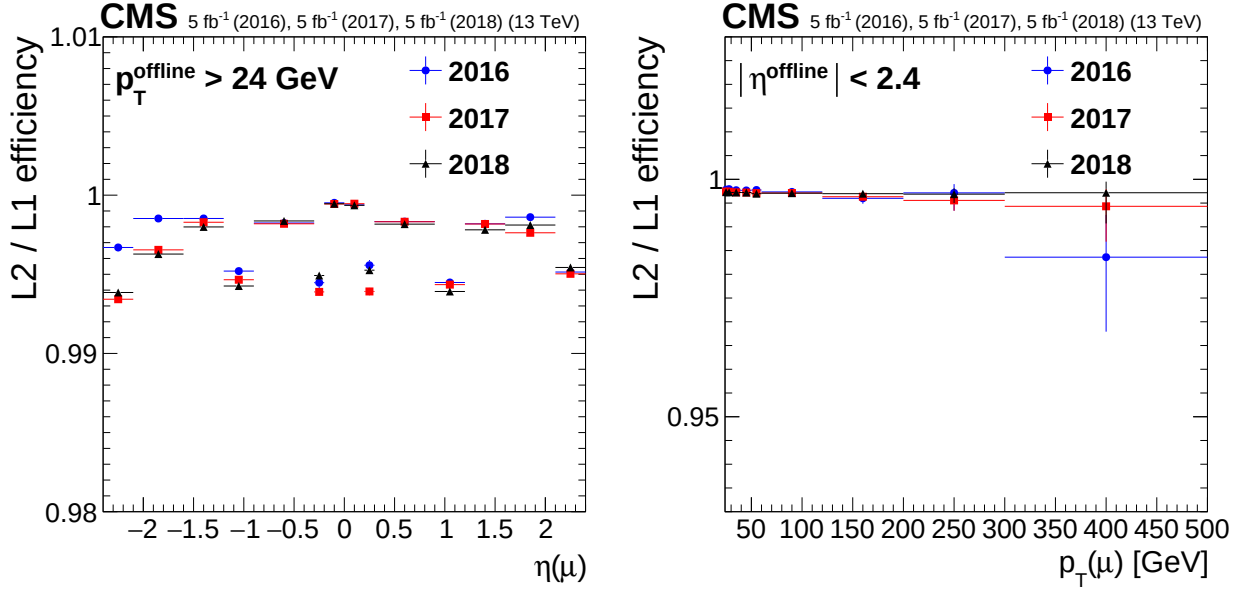


Figure 3.1. L2/L1 efficiencies in the Run 2 data [106].

Trajectory seeds in L3 reconstruction.

As mentioned above, the Kalman filter algorithm requires an initial estimate of the parameters at some point of the trajectory. This estimate is referred to as *seed*. The starting point for seed generation is the L2 TSOS at the innermost muon station.

The outside-in L3 reconstruction in Run 2 used two types of seeds: *hitless* and *hitbased*. For Run 3, a new type of seeds, *hitbased doublets* has been developed (see Section 3.3.2). The ideas behind different types of seeds are illustrated in Figure 3.2.

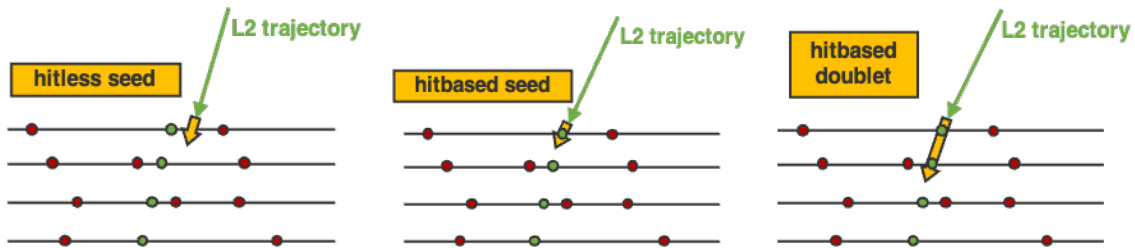


Figure 3.2. Left: a *hitless seed* is created by propagating the L2 trajectory to a detector layer. Center: a *hitbased seed* is created by updating the propagated trajectory with the most compatible hit. Right: a *hitbased doublet* is created by updating the trajectory with two hits on consecutive layers.

A hitless (HL) seed is generated by propagating a L2 TSOS to a given tracker layer, and using the parameters and uncertainties of that TSOS. Hitless seeds can be created either starting from a L2 TSOS updated at the interaction point (IP), or, alternatively, from the TSOS at the innermost muon station without the IP constraint. In both cases, the TSOS is propagated to the outer tracker surface, and then to the inner-more layers (hence the name “outside-in”).

A hitbased (HB) seed is created in a similar way, but when a TSOS is propagated to a layer, it is updated with the most compatible hit on that layer; the figure of merit for the compatibility of hits is the normalized chi-squared test statistics χ^2 [107], with a requirement that $\chi^2 < 40$. In the trajectory building stage, the hit used to create a HB seed is included in the trajectory. The efficiency achieved with hitbased seeds is generally lower than that for HL seeds (see Figure 3.3), due to a significant amount of pile-up hits in the tracker. A trajectory reconstructed from a HB seed may be diverted from its true position if it is

updated with a pile-up hit, which does not happen to HL-seeded trajectories at this stage of the reconstruction.

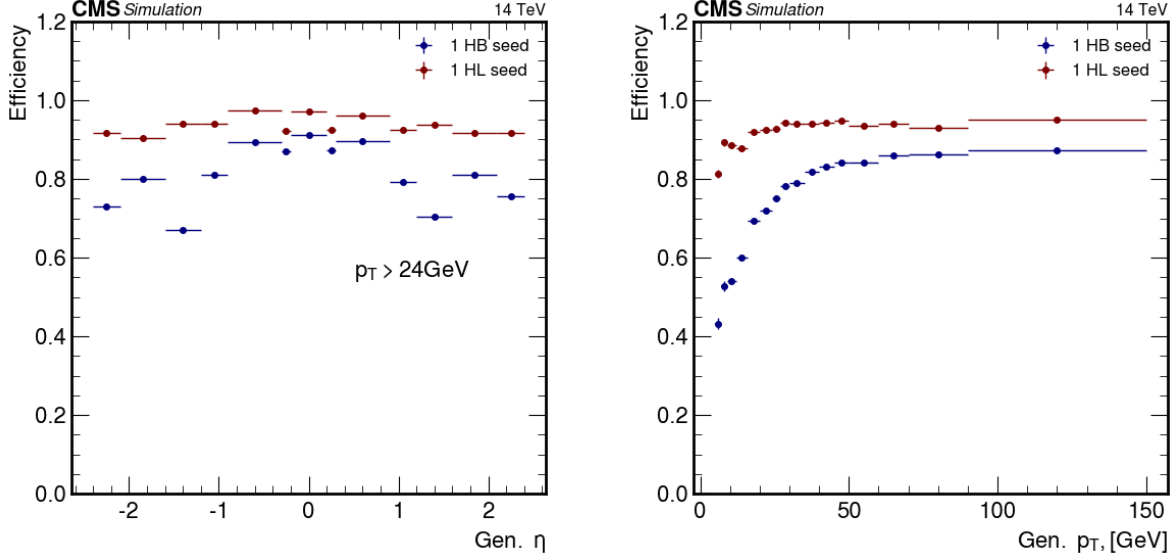


Figure 3.3. L3(OI)/L2 efficiency for trajectory building seeded with a single HB or single HL seed.

The pile-up hits also affect the trajectory building itself. As mentioned earlier, the Kalman filter algorithm searches for compatible hits at each iteration and uses them to create trajectory candidates. In the presence of pile-up it is possible that at a certain step all of the most compatible hits will be from pile-up, leading to eventual failure of trajectory building. This effect can be partially suppressed by creating multiple seeds on different tracker layers. Even though hitless seeds created from the same L2 muon carry the same information, the probability of successful L3 reconstruction for them is slightly different due to the different distribution of pile-up hits at each layer. Figure 3.4 shows the L3 efficiency as a function of pile-up (characterized by the number of reconstructed vertices in the event $nVtx$); it can be observed that creating seeds on more layers leads to a larger increase in efficiency in events with more pile-up. The scaling of the efficiency with the number of HL seeds as a function of other variables is shown in Figure 3.5.

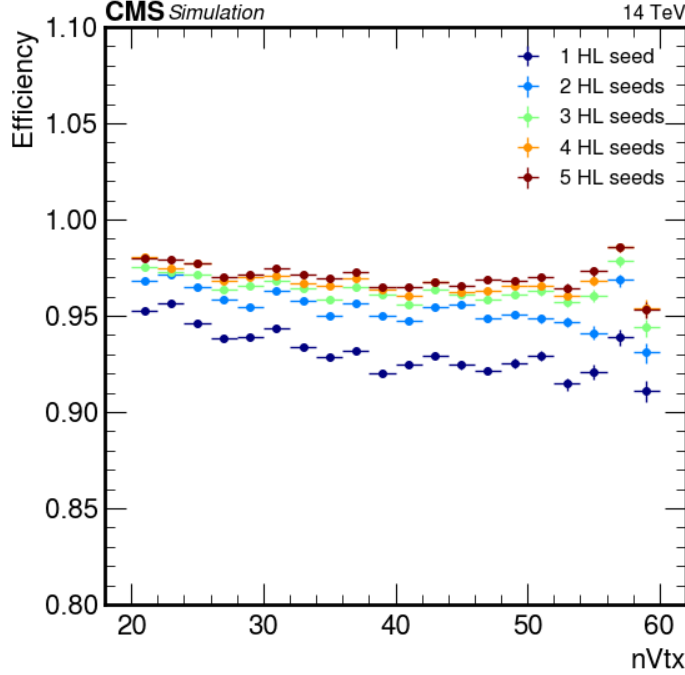


Figure 3.4. L3(OI)/L2 efficiency as a function of pile-up. The increase in efficiency from additional seeds is correlated with the number of pile-up vertices.

TSOS error scale factors.

When the CTF algorithm is initialized using a HL seed, the search window for the compatible hits is defined by the uncertainties on the trajectory parameters, which are stored in a *covariance matrix* associated with the seed. However, these uncertainties do not always correspond to the optimal size of the search window needed for successful performance of the trajectory builder. Unlike HB seeds, the HL seeds are not based on any tracker information, and therefore their precision can not be improved by the inclusion of tracker hits. Instead, the possible inaccuracy of the initial estimate can be compensated by enlarging the uncertainties, which allows the CTF algorithm to search for compatible hits in wider search windows. This is implemented by introducing *TSOS error scale factors*, which in the Run 2 strategy were defined in bins by η and p_T , and ranged from 2 to 10 (see Table 3.1).

While the enlarged uncertainties help to find compatible hits in case of an inaccurate initial estimate, it does *not* mean that any trajectory can be successfully reconstructed from

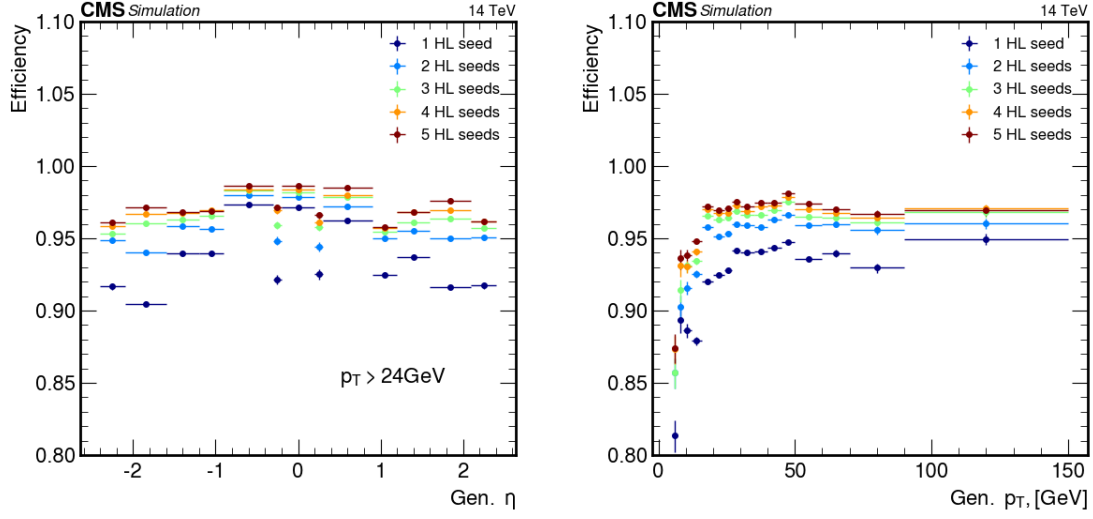


Figure 3.5. L3(OI)/L2 efficiency for trajectory building seeded with various numbers of hitless seeds.

Table 3.1. Error scale factors used in Run 2 [108].

$ \eta $	$p_T : 0 - 13 \text{ GeV}$	$p_T : 13 - 30 \text{ GeV}$	$p_T : 30 - 70 \text{ GeV}$	$p_T \geq 70 \text{ GeV}$
0 – 0.2	3	5	2	10
0.2 – 0.3	3	5	7	10
0.3 – 1.0	3	5	2	10
1.0 – 1.2	3	4	3	7
1.2 – 1.4	3	4	3	7
1.4 – 1.6	3	4	3	7
>1.6	3	5	10	10

a HL seed as long as the error scale factor is large enough. The CTF algorithm is designed to avoid high costs in timing, therefore it aborts if failing to build a trajectory after trying several (by default, five) compatible hits on each iteration of the Kalman filter. When the search window is enlarged, it includes more pile-up hits, so the CTF algorithm that initiates trajectory building for only five of the compatible hits may exhaust all of the attempts before finding the “true” hit.

Therefore, the optimal scale factors for each seed are finite, and empirical observations show that the optimal values for the scale factors are of the order of 1–10.

Seeding strategy.

For a given seed created from an L2 muon in the outer tracker, it is not guaranteed that the Kalman filter will succeed in the L3 reconstruction. The main reason for this is a substantial amount of tracker hits that are not produced by the considered muon (such as pile-up or detector noise). At each iteration of the Kalman filter, the search window for finding compatible hits is determined by the parameters and uncertainties obtained at the previous iteration. If at some iteration the trajectory was updated with a pile-up hit, the new trajectory may get diverted so much that the “true” muon hits are not compatible with it anymore, and trajectory reconstruction for the chosen seed will eventually fail.

In order to maximize the chance of success, multiple HL and/or HB seeds on different tracker layers can be generated for each L2 muon. A particular placement of seeds is referred to as *seeding strategy*.

In Run 2, multiple seeds of the same type could be created only on multiple consecutive outer tracker layers; strategies with other seed placements (e.g. with seeds on non-consecutive layers) were not considered. Therefore, a seeding strategy in general could be described by two parameters, the number of HB and the number of HL seeds. The details of the Run 2 seeding strategy are listed in the next section.

The new developments for Run 3, described in the following sections, use a similar approach to the seed placement, creating the seeds only on consecutive outer layers, unless stated otherwise.

3.2 L3 muon reconstruction in Run 2

The muon HLT in the Run 2 setup was implemented via an iterative algorithm [108] containing three sequences in increasing order of complexity:

1. Outside-in reconstruction seeded by L2 muons (OI)
2. Inside-out reconstruction seeded by L2 muons (IOFromL2)
3. Inside-out reconstruction seeded by L1 muons (IOFromL1)

The inside-out reconstruction seeded by L2 muons is only performed for those L2 muons for which no tracker track was reconstructed in the outside-in step, while the inside-out from L1 step is seeded by all L1 muons. The resulting track collections from each of the three steps are merged and matched with the information from the muon system, producing the final L3 muon objects. Loose muon ID criteria are then applied to filter the resulting muon collection to suppress misreconstructed muons and control the trigger rate. The efficiency of all three steps of the Run 2 strategy is shown in Figure 3.6.

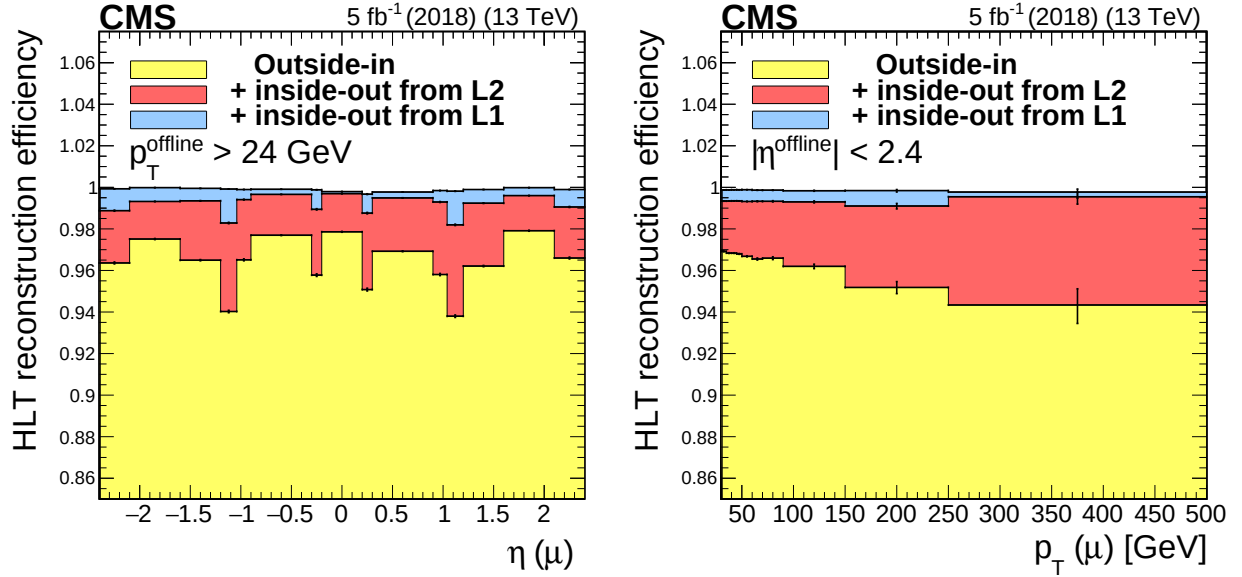


Figure 3.6. L3/L1 efficiencies in Run 2 [106].

3.2.1 Outside-in seeding in Run 2

The seeding strategy in Run 2 was chosen based on a manual optimization of the strategy parameters (number of seeds of each type), as well as some fine-tuning aimed to boost the performance in various regions of the phase space. The definition of the Run 2 outside-in seeding strategy is summarized below:

- The baseline strategy contained 1 HB seed and 5 HL seeds per L2 muon.
- In the barrel region ($|\eta| < 1$), the HB seeds were disabled.

- If a trajectory intersected the tracker layers in both barrel (TOB) and endcap (TEC), the seeds of each type were counted in TOB and TEC independently, thus allowing to create up to 2 HB and 10 HL seeds in the barrel-endcap overlap region. In practice, the creation of the maximum allowed number of seeds in the overlap is never observed; the total number of seeds in the overlap is at most 8.
- In the forward region ($|\eta| > 2.1$) and for muons with less than 20 valid hits in the muon system, the number of hitless seeds was doubled. The additional hitless seeds were created using an alternative L2 trajectory, not constrained at the interaction point. This lead to creating 11 seeds (1 HB + 10 HL) in the forward region and for muons with only a small number of valid hits.
- The errors associated with hitless seeds were rescaled by factors ranging from 2 to 10, depending on p_T and η .

As a result, the number of seeds per L2 muon in Run 2 ranged from 5 to 11, depending on the L2 muon quality and location in the detector. In the overlap region, where the barrel and endcap layers were considered independently, the total number of seeds created for a single L2 muon could be as high as 15; however, the number of such events was low. The distribution of number of seeds for the Run 2 strategy is shown in Figure 3.7.

3.3 Outside-in seeding developments for Run 3

The overall design of muon HLT in Run 3 (Figure 3.8) is similar to that used in Run 2, featuring the three steps (OI, IOFromL2, IOFromL1) described in Section 3.2. However, the seeding algorithms used in the OI and IO steps have been updated.

The changes to the outside-in reconstruction for Run 3 (seed types, treatment of L2 trajectories, etc.) are described in this section. The optimization of the OI seeding strategy is described in the next section.

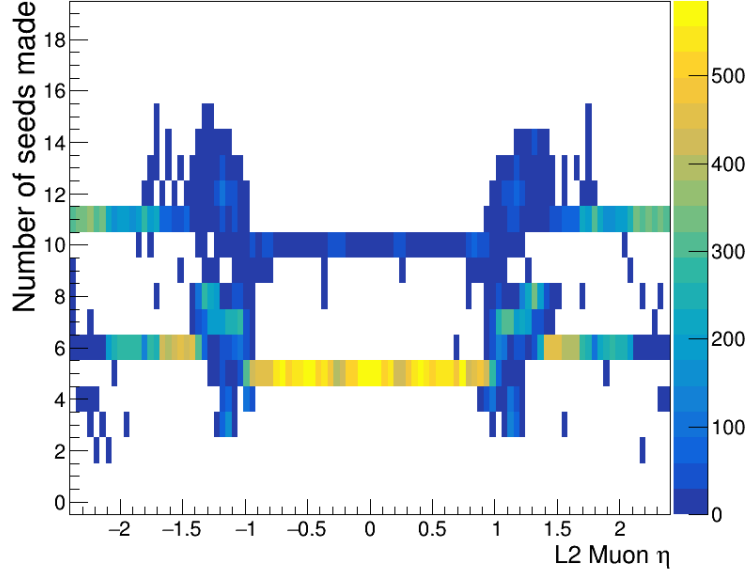


Figure 3.7. Total number of seeds created in the Run 2 strategy, as a function of η .

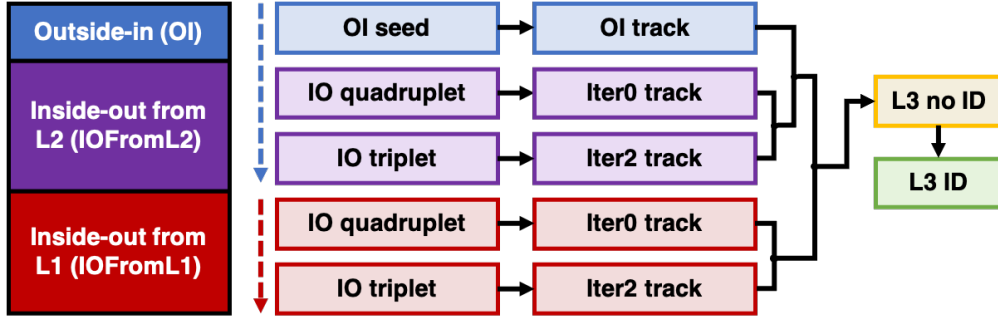


Figure 3.8. Proposed muon HLT design for Run 3.

3.3.1 Changes to the ad hoc features remaining from the Run 2 strategy

In the Run 2 strategy, the efficiency in the forward η region and for low-quality L2 muons was boosted by considering two L2 trajectories for the same muon: one trajectory reconstructed only using the information from the muon system, and another one updated at the interaction point. Using this additional L2 information has indeed lead to improved efficiency, but achieving it by doubling the number of hitless seeds came with a large cost in terms of CPU time. This approach is therefore discarded for Run 3, given that the main objective of the Run 3 developments is to improve the timing of the OI sequence.

In the new approach, the seeds created from the two L2 trajectories are considered as separate types of seeds. The trajectories and corresponding seeds are labeled “MuS” (L2 trajectory containing only the hits from muon system) and “IP” (L2 trajectory updated at the interaction point). The seeds of the new types can be generated independently from each other, allowing a more flexible control over the number of seeds.

Another feature of the Run 2 approach was a veto on hitbased seeds in the $|\eta| < 1$ region, which was enabled in an attempt to improve the timing. In the new developments, the veto has been removed in favor of more flexible and automatized ways to control the efficiency and timing of the seeding strategies (see Section 3.4).

3.3.2 Hitbased doublet seeds

Hitbased seeds in the Run 2 algorithm were created by choosing the hit most compatible with the L2 trajectory. The compatibility was quantified by χ^2 for the trajectory updated with the new hit, provided that $\chi^2 < 40$. The TSOS updated with the most compatible hit was used as a starting point for the L3 trajectory building sequence.

The problem with this approach appeared when the L2 trajectory propagated to the outer tracker diverted too much from the real muon trajectory. In this scenario, it was possible that at the outermost tracker layer a pile-up hit was more compatible with an L2 trajectory than a “real” muon hit. A L2 trajectory updated with a pile-up hit would then give a wrong estimate of the parameters, leading to the failure of the L3 reconstruction.

It has been observed that the Kalman filter algorithm seeded by a TSOS with wrong parameters tends to fail in the first iteration. In other words, it is unlikely that the tracker would contain two or more pile-up hits on multiple consecutive outer tracker layers, such that those hits would all be compatible with a given L2 trajectory.

An improvement to the hitbased seed definition has been developed based on this observation. As illustrated in Figure 3.9, in the new implementation a L2 trajectory is only updated with a tracker hit if there is another compatible hit on the adjacent tracker layer; otherwise the hit is skipped and the next most compatible hit is considered. If two hits compatible with the L2 trajectory are found, they are both used to update the trajectory

and create a seed for the L3 trajectory builder. Essentially, this approach implements the first step of the Kalman filter at the seed generation stage, which allows to reject bad hits early and with only a small timing cost.

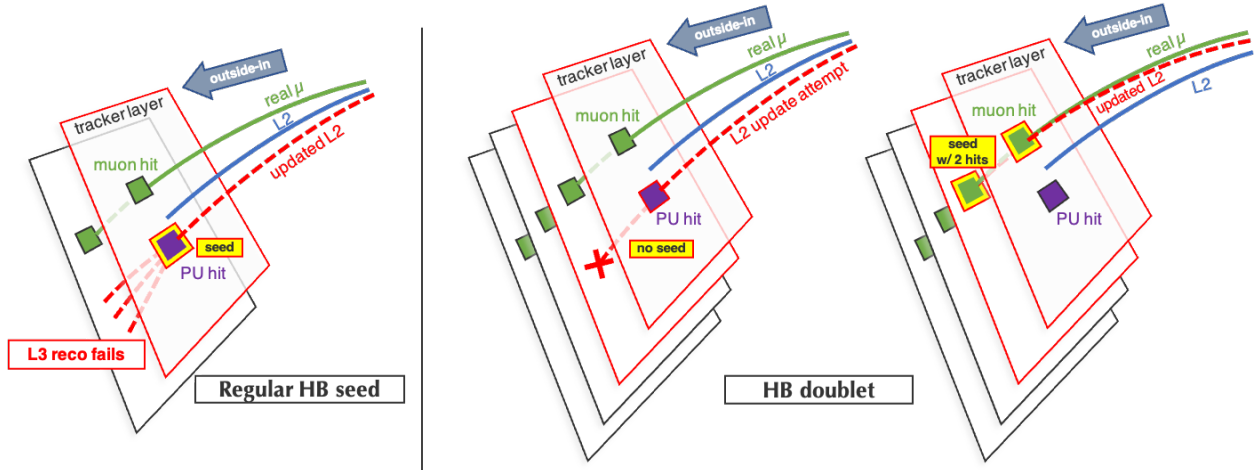


Figure 3.9. Left: a regular HB seed constructed from the most compatible hit which can be from pile-up. Center: a seed is not generated if there are no compatible hits after update. Right: a HB-doublet seed is generated if there are compatible hits on two layers.

The new type of hitbased seeds is referred to as *hitbased doublets*. A single HB-doublet uses approximately the same information as two regular HB seeds, and gives a comparable performance in terms of efficiency (see Figure 3.10), but the timing is reduced, as shown in Figure 3.11, since the full trajectory building using the Kalman filter is only run once in case of the doublet, and twice in case of the two regular seeds.

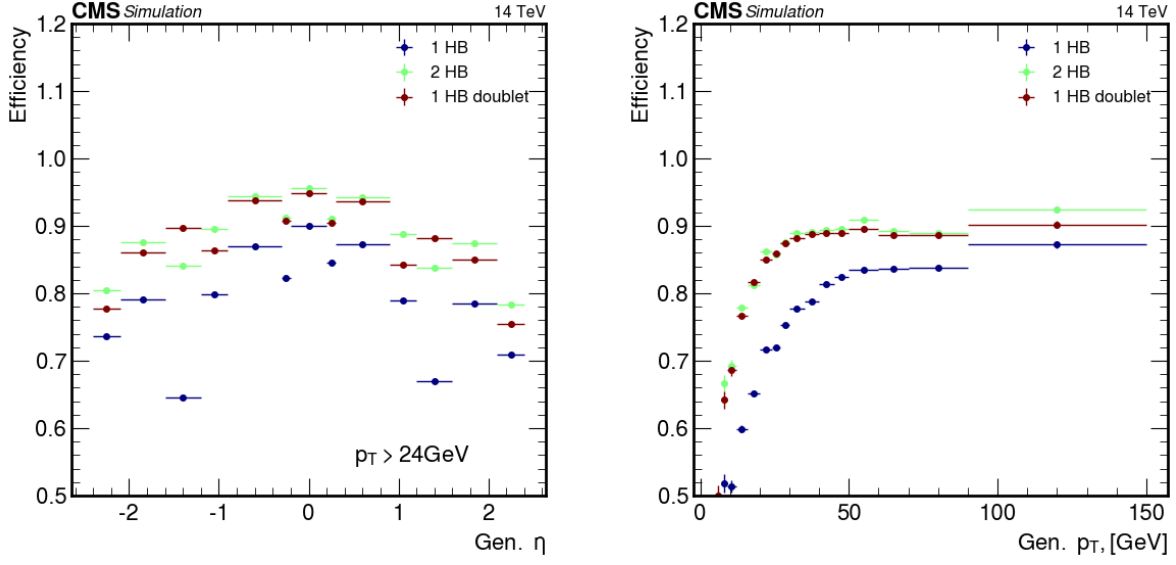


Figure 3.10. Efficiency with a single HB-doublet seed is similar to that of two regular HB seeds.

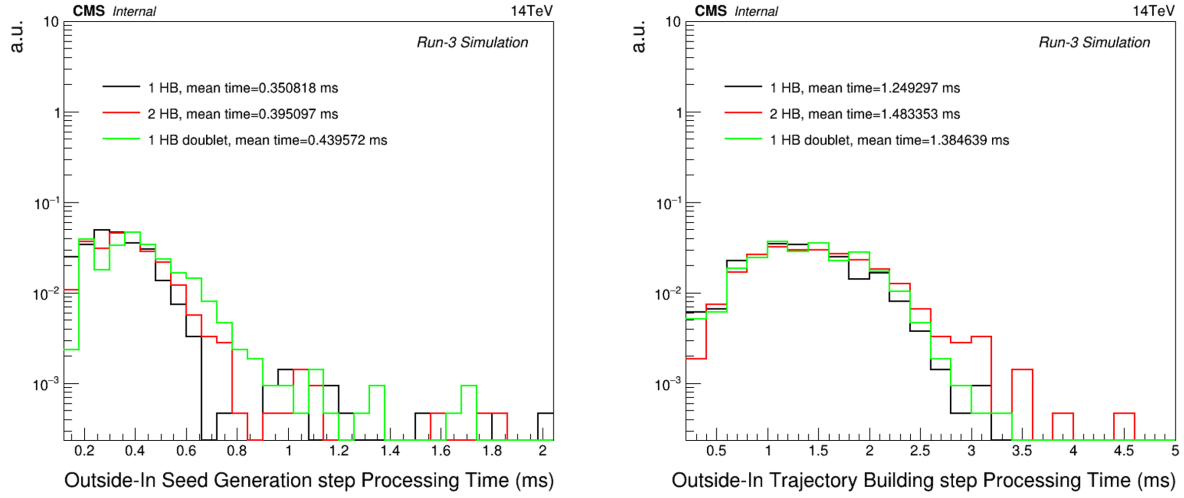


Figure 3.11. Processing time of seed generation and trajectory building for regular and doublet-like HB seeds.

3.4 DNN-based seeding strategy optimization

As described in Section 3.2, the OI seeding strategy developed for Run 2 featured a certain flexibility: the input L2 muons were categorized according to their pseudorapidity and the number of valid hits in the muon system, and a separate seeding strategy was defined for each category. The new approach aims to expand this flexibility by defining the seeding strategy independently *for each L2 muon*, rather than, or in addition to considering muons in broad categories.

3.4.1 Problem definition

If L2 muons are considered independently, the optimization of the outside-in seeding can be parametrized in the following way:

- A set of input features $\bar{x} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N\}$ characterizing each of the N L2 muons in a given dataset.
- A set of parameters $\bar{\theta}_i$ characterizing the seed generation strategy for i -th L2 muon¹. These parameters include the number of seeds of each type, as well as the value of TSOS scale factors for HL seeds.
- Figures of merit that quantify the performance of the method on a given dataset D : algorithmic efficiency $\epsilon(D, \bar{\theta})$ and the time for trajectory building $T(D, \bar{\theta})$.

The optimization problem is then defined as follows: for each set of features $\bar{x}_i$ corresponding to the i -th L2 muon in a given dataset D , obtain a set of strategy parameters $\bar{\theta}_i^*$, such that the figures of merit $\epsilon(D, \bar{\theta})$ and $T(D, \bar{\theta})$ are optimized.

In the developments for Run 3, the trajectory parameters are defined as follows:

$$\bar{\theta} = \{\text{nHBd}, \text{nHL(IP)}, \text{nHL(MuS)}, \text{SF}\},$$

¹↑In this section, the indexing of the variables is defined as follows: lower indices $i = \{1, 2, \dots, N\}$ enumerate L2 muons in the dataset, and upper indices $j = \{1, 2, \dots, K\}$ enumerate considered seeding strategies.

where n_{HBd} is the number of HB-doublet seeds (as described in Section 3.3.2), $n_{HL}(\text{MuS})$ is the number of HL seeds from L2 trajectory based only on the muon system information, $n_{HL}(\text{IP})$ is the number of HL seeds from L2 trajectory updated at interaction point, and SF is the error scale factor used for HL seeds — the same for both $HL(\text{IP})$ and $HL(\text{MuS})$. In principle, different scale factors can be used for different types of seeds, but it was not done in order to avoid increasing the dimensionality of $\bar{\theta}$.

3.4.2 A new optimization approach

The solution of the optimization problem is implemented using a deep neural network (DNN) — a machine learning method, described in more detail in Appendix B. Given a set of K seeding strategies $\bar{\Theta} = \{\bar{\theta}^1, \bar{\theta}^2, \dots, \bar{\theta}^K\}$, the network is trained to select the strategy that is more likely to lead to the successful L3 reconstruction for a given L2 muon, which is parameterized by a set of input features $\bar{x}$. The schematic depiction of the network design is shown in Figure 3.12.

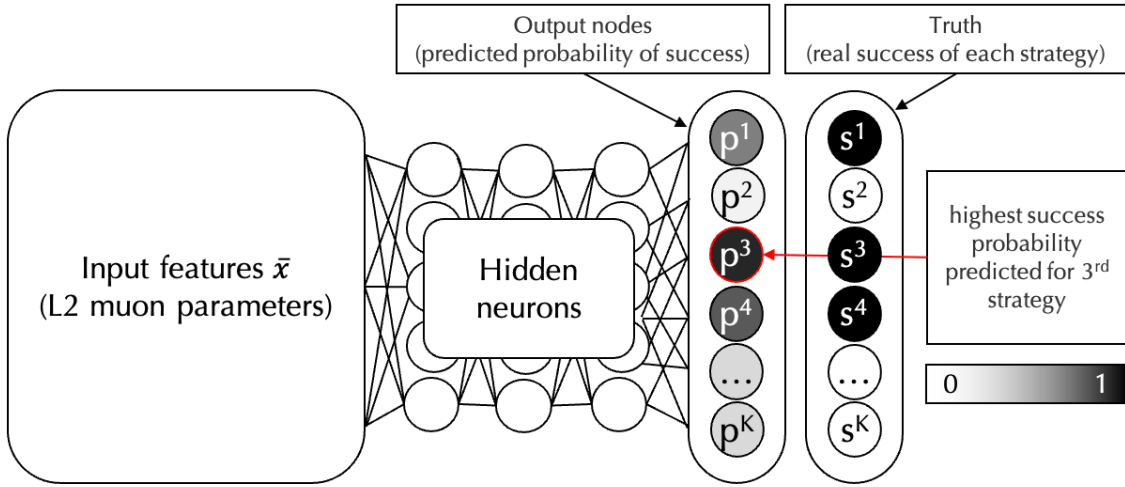


Figure 3.12. DNN classifier used to predict the seeding strategy with highest probability of success.

In order to train the classifier to choose between the seeding strategies in $\bar{\Theta}$, the training dataset is constructed as follows:

1. The L3 reconstruction algorithm is run K times on the same set of input L2 muons (once for each seeding strategy).
2. An attempt is made to match each of the reconstructed L3 muons to the generated muons in the same event, in order to determine whether the L3 reconstruction was successful. The successfully reconstructed L3 muons are matched to the respective L2 muons. In both cases, the matching is performed by checking if the angular separation between the two objects is $\Delta R < 0.3$, where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.
3. After the matching, there is a set of K numbers $\bar{S}_i = \{s_i^1, s_i^2, \dots, s_i^K\}$ for each of the input L2 muons, which encodes the success of each seeding strategy for that L2 muon: $s_i^j = 1$ if j -th strategy for i -th muon led to a successful L3 reconstruction, $s_i^j = 0$ otherwise.
4. In addition to the s_i^j values, the training dataset includes the set of input features $\bar{x}_i$ for each muon.

If a given set of strategies $\bar{\Theta} = \{\bar{\theta}^j, j = 1 \dots K\}$ is considered, the optimization problem can be reformulated in terms of multi-class classification: a L2 muon “belongs” to the j -th “class” if the seeding strategy $\bar{\theta}^j$ associated with that class leads to successful L3 reconstruction; each muon can belong to multiple classes, since multiple strategies can succeed for the same muon.

3.4.3 Input features

The input features for the classifier are chosen to use as much available information about the L2 muons as possible:

- Global p_T , η and ϕ of the L2 muon
- Number of valid hits in the muon system

- p_T , η and ϕ of the TSOS defined at the IP
- p_T , η and ϕ of the TSOS defined at the innermost muon station
- Diagonal elements of the 5×5 covariance matrices defined at the IP and at the innermost muon station

The correlations between the considered input features are shown in Figure 3.13. Most of the variables are uncorrelated with each other, with several exceptions:

- The kinematic parameters of the TSOS at interaction point are 100% correlated with the corresponding global kinematics of the L2 muon. Therefore, they are removed from the set of input features in the training dataset.
- Uncertainties on trajectory parameters (variables labeled `err*`) for both TSOS at the interaction point and the innermost muon station are correlated between each other. All of these variables are used in the training and no additional steps are taken to decorrelate them.

For any given seeding strategy, the success of the L3 reconstruction is correlated with the input features. As an example, the distributions of the input features for events passing and failing L3 reconstruction seeded by a single HL(IP) seed are shown in Figure 3.14. Such correlations are slightly different for various seeding strategies (see Figure 3.15), and the DNN classifier is trained to learn these subtle differences.

3.4.4 DNN architecture and training settings

The network contains two fully connected hidden layers with 512 neurons each, as well as the output layer with K nodes. The hidden layers are subjected to tanh activation functions to constrain the values of the weights, and the sigmoid function [109] is applied to the output nodes, which ensures that the values of the outputs are between 0 and 1.

The output of the network for a given muon yields a set of values $\bar{P}_i = \{p_i^1, p_i^2, \dots, p_i^K\}$, $0 \leq p_i^j \leq 1$, which can be interpreted as predicted probabilities of success for each seeding strategy. The predicted probabilities $\bar{P}_i$ are compared to the “truth” $\bar{S}_i$, and the binary

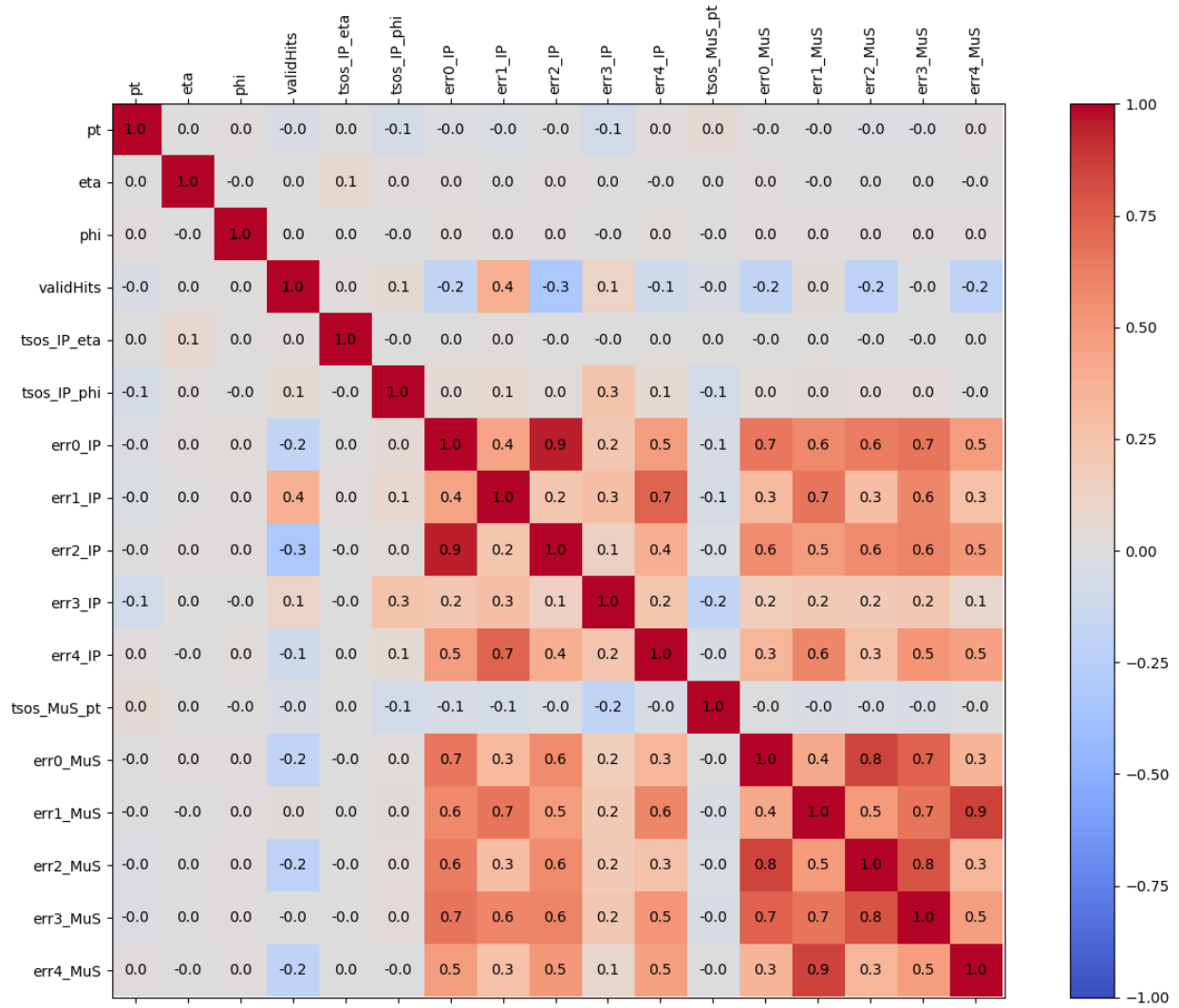


Figure 3.13. Correlations between the DNN input features.

cross-entropy loss function is constructed. During the training process, the loss function is minimized using the Adam optimization method [110]. At each training iteration, 20% of weights between the hidden layers are removed in order to improve the stability of the training and prevent overfitting.

The trained network can be applied to a new L2 muon, producing a set of predictions $\bar{P}^*$. The optimal strategy $\bar{\theta}^*$ is then selected out of $\bar{\Theta}$ by picking the strategy with index

$$j^* = \operatorname{argmax}_j \bar{P}^*. \quad (3.2)$$

An example of the strategy prediction by the DNN classifier is shown in Figure 3.16, where only the strategies with two seeds were considered.

Since each seed used in the L3 reconstruction covers a slightly different region of the phase space, the strategies with a higher number of seeds overall tend to show better efficiency. However, the timing of the trajectory building scales almost linearly with the total number of created seeds. Therefore, the method that optimizes efficiency of a seeding strategy does not necessarily minimize the timing; the DNN optimization tends to pick the strategies with the higher number of seeds, which leads to slower trajectory building.

To overcome this issue, the training was done using sets of strategies with a fixed total number of seeds. The performance of the strategies with the total number of seeds ranging from 1 to 7 is shown in Figure 3.17. In this case, the error scale factors were not optimized.

It can be seen that the efficiency scales differently with the number of seeds in the barrel and endcap regions. For the final setup, the 5-seed strategy was chosen in the barrel ($|\eta| < 1$) and the 7-seed strategy was used in the endcaps ($|\eta| > 1$).

The training was then repeated with scale factor optimization included. The evolution of training and validation loss for the two trained network is presented in Figure 3.18, showing no signs of overfitting.

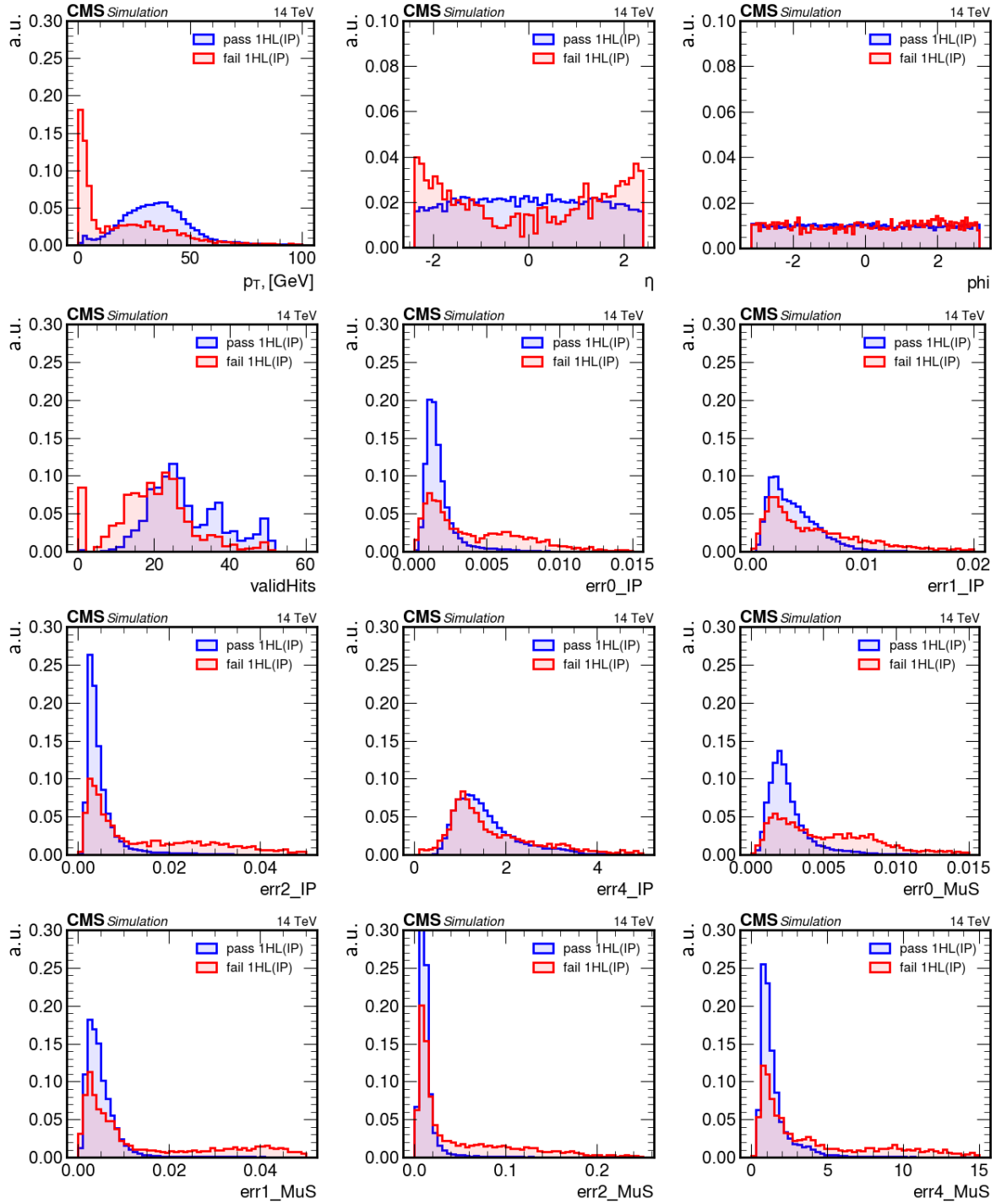


Figure 3.14. Distribution of some L2 parameters for events passing and failing L3 reconstruction seeded by a single HL(IP) seed. Variables labeled “err” are square roots of diagonal elements of covariance matrices corresponding to the L2 trajectories, i.e. uncertainties on global trajectory parameters.

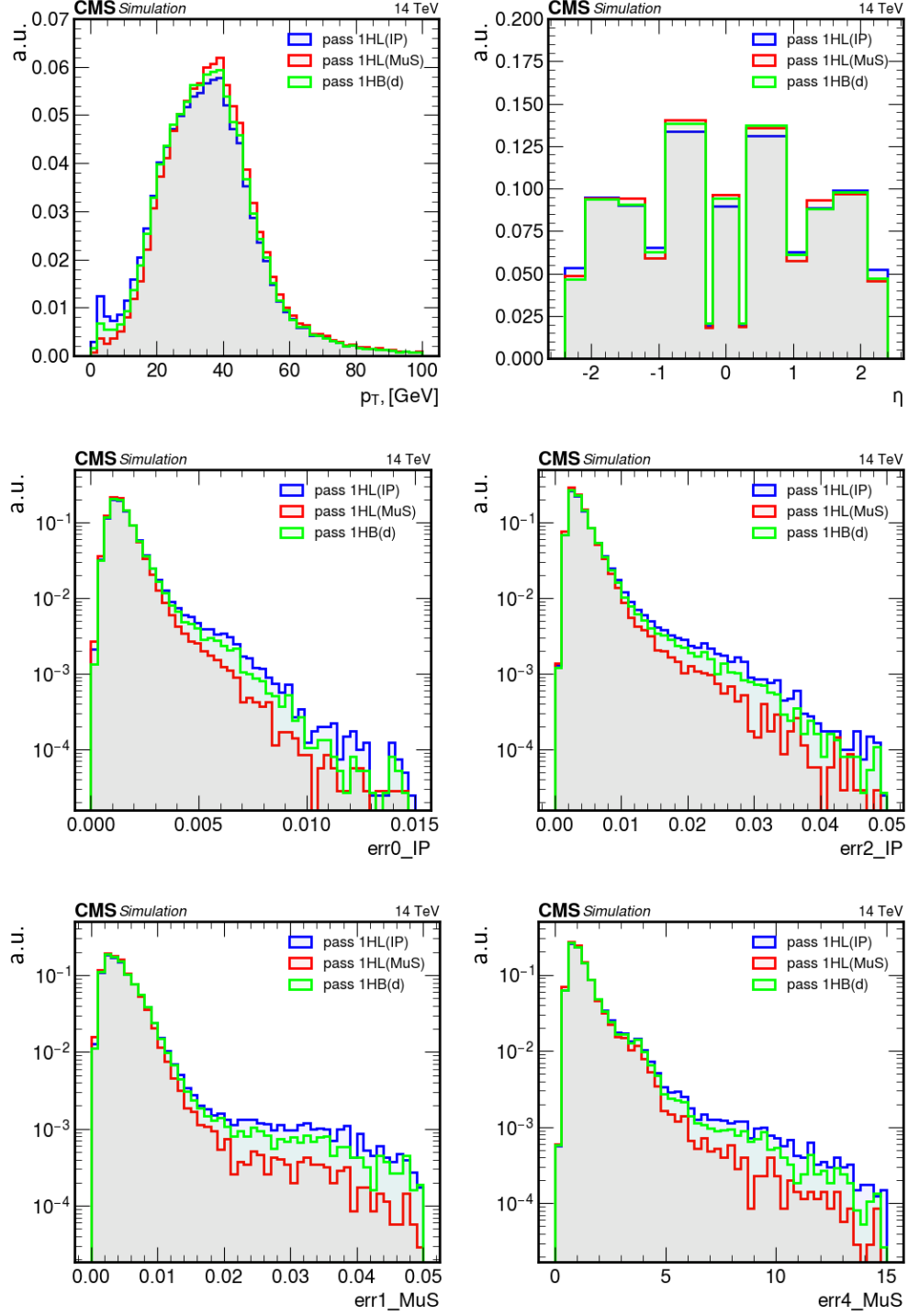


Figure 3.15. Distribution of some L2 parameters for events passing L3 reconstruction seeded by different types of seeds. The DNN classifier is trained to learn the differences between these distributions.

3.4.5 Performance

The L3 efficiency when using the strategy optimized with a DNN classifier is compared to that for the default Run 2 strategy in Figure 3.19. The efficiency obtained using the DNN classifier is the same or higher than that for the Run 2 strategy in almost all regions of phase space, while the number of created seeds is smaller on average.

Timing comparison of these options is shown in Figure 3.20. Evaluation of the classifier at runtime adds ≈ 0.3 ms per muon to the seed generation time, while the trajectory building time is on average 3 ms (23%) faster compared to that for the Run 2 strategy. Therefore, the overall timing is decreased.

The novel DNN-based seed generation algorithm for the outside-in L3 muon reconstruction has been deployed to the CMS detector after the beginning of the LHC Run 3 in July 2022. For the final implementation, the model was re-trained with low- p_T muons ($p_T < 24$ GeV) added to the training dataset, and the model architecture was once again optimized using the KERASTUNER package [111].

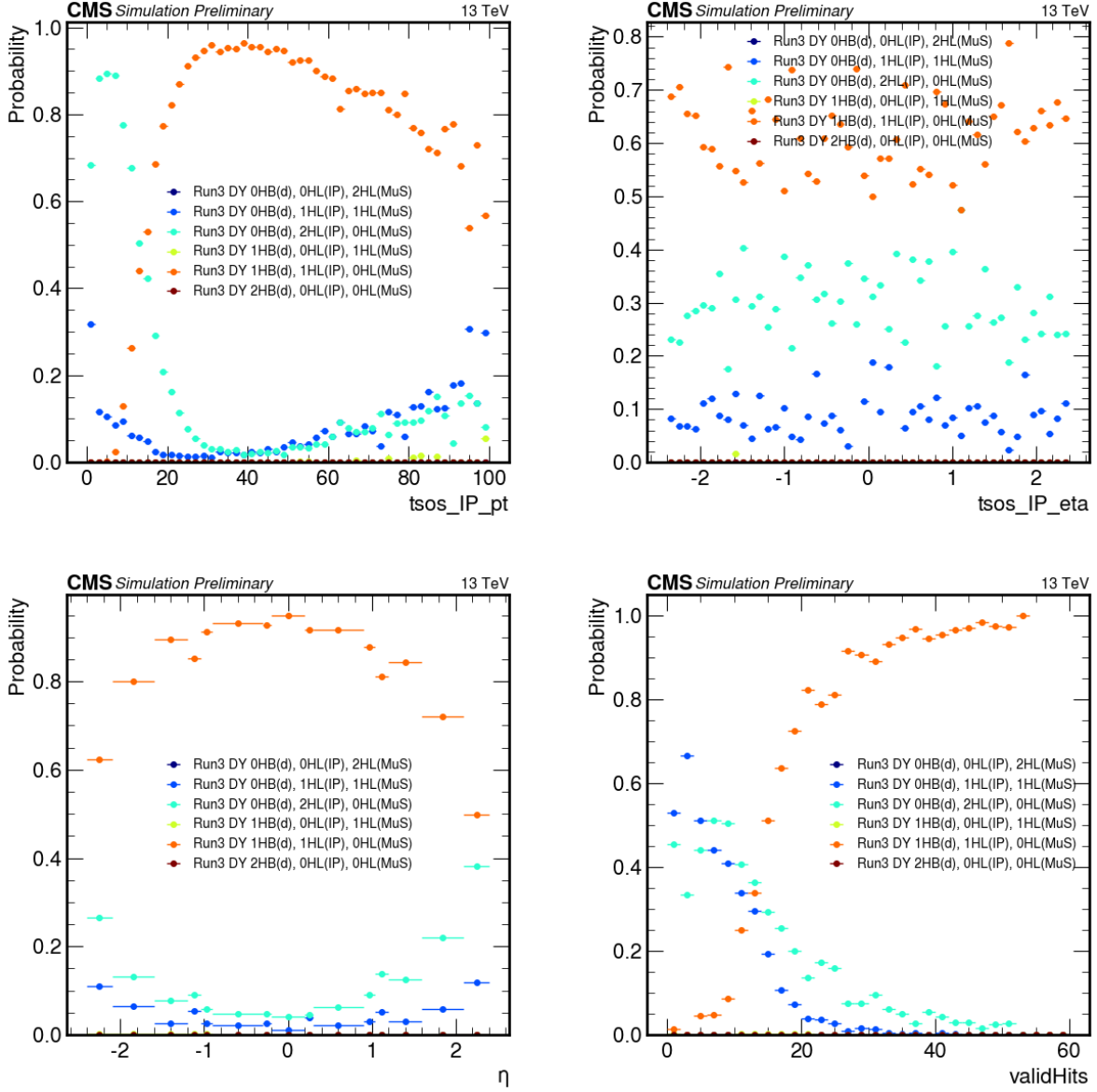


Figure 3.16. Example of seeding strategies predicted by the DNN classifier trained to choose between 2-seed strategies without error scale factor optimization. In this example, the strategy with 1 HB-doublet and 1 HL(IP) seed is preferred for high-quality muons, while strategies with only hitless seeds are preferred for lower quality muons.

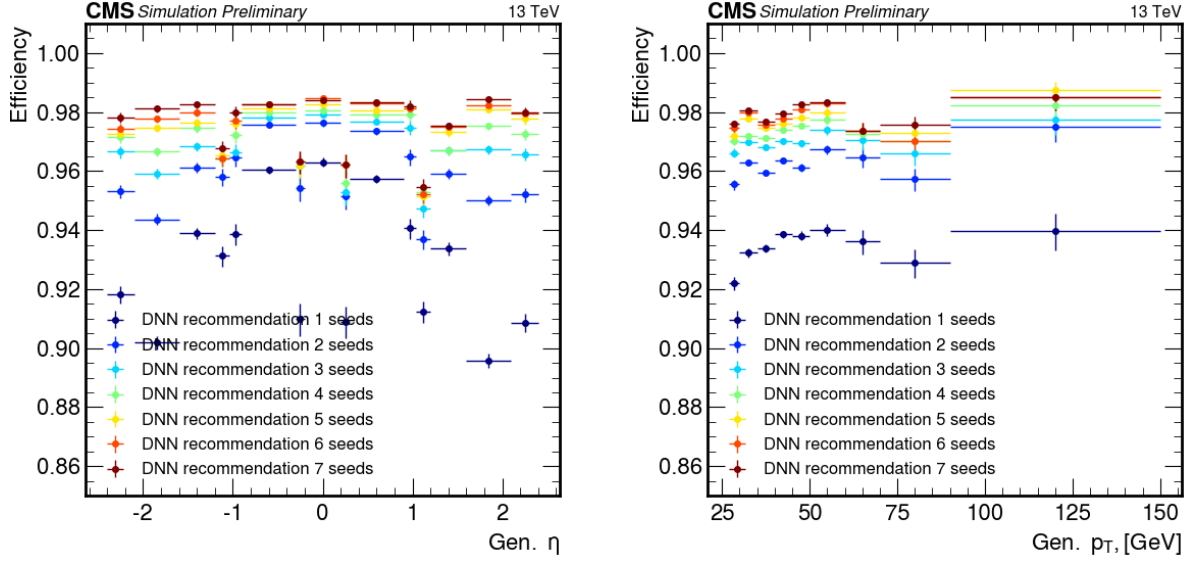


Figure 3.17. L3 efficiency of optimized strategies with fixed total number of seeds ranging from 1 to 7.

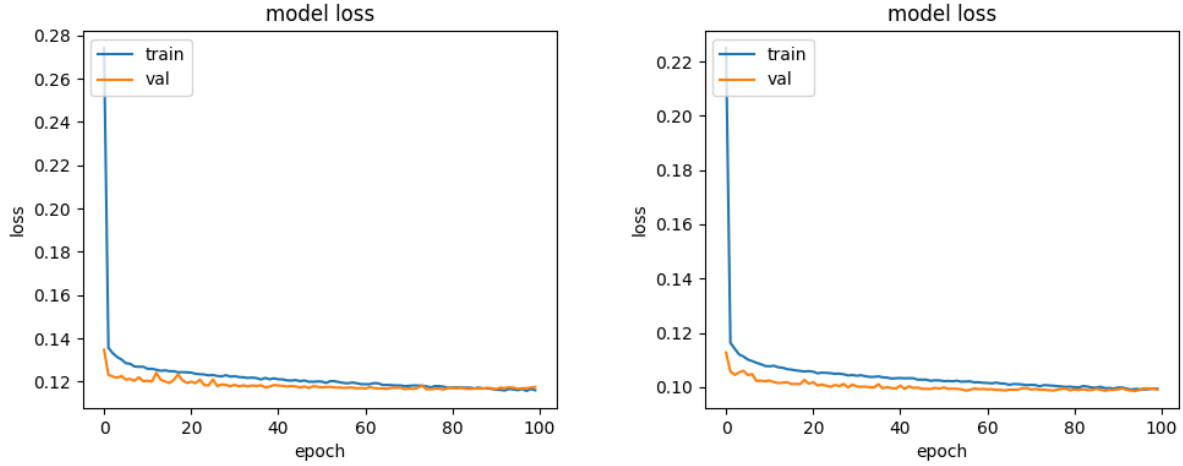


Figure 3.18. Training and validation loss for DNN selecting between 5-seed strategies, used in the barrel region (left), and for DNN selecting between 7-seed strategies, used in the overlap region (right).

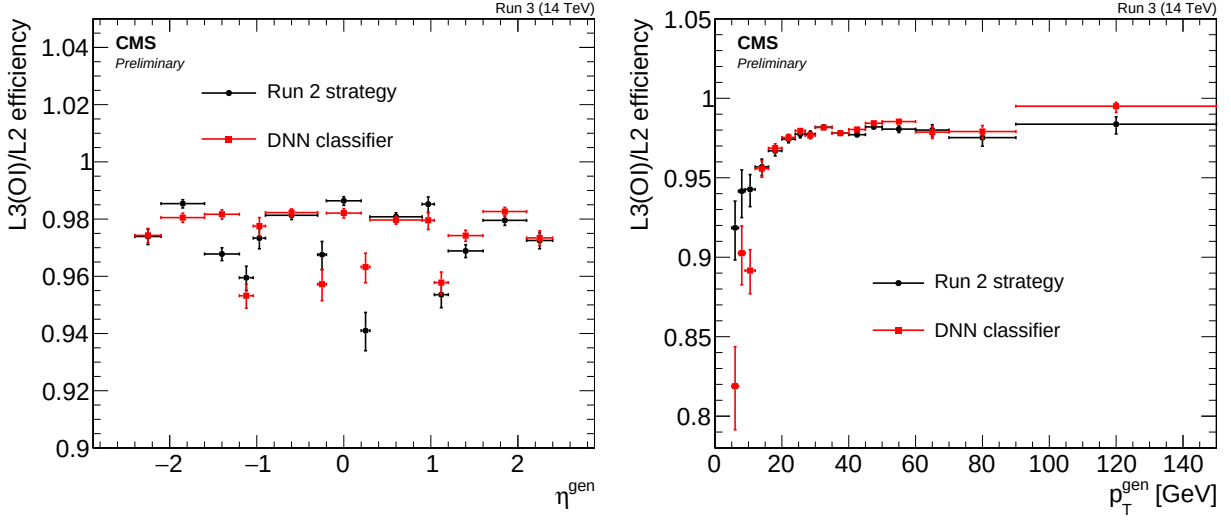


Figure 3.19. L3 efficiency for the strategy optimized using DNN classifier, compared to the default Run 2 strategy. The efficiency is measured using a Drell–Yan dataset.

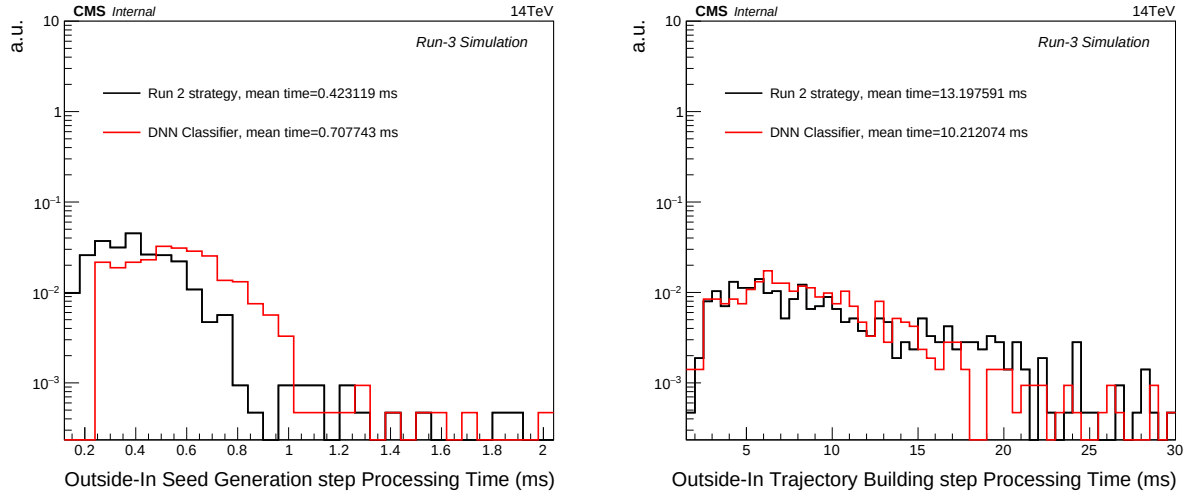


Figure 3.20. Seed generation (left) and trajectory building (right) time for the DNN classifier compared to the Run 2 approach. The timing is measured using a Drell–Yan dataset.

4. FIRST EVIDENCE OF THE $H \rightarrow \mu\mu$ DECAY

The decay of the Higgs boson into a pair of muons ($H \rightarrow \mu\mu$) provides a direct way to establish the Higgs coupling to muons, which is one of the free parameters in the Standard Model. The branching ratio predicted by the SM for the $H \rightarrow \mu\mu$ decay is small in comparison to other Higgs decay modes (see Section 1.3.2), therefore the rate of such decays at the LHC is expected to be relatively low.

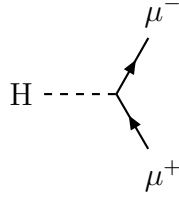


Figure 4.1. Feynman diagram of the $H \rightarrow \mu\mu$ decay.

Achieving a statistically significant measurement of the $H \rightarrow \mu\mu$ cross section is a challenging task that requires an accurate modeling of the SM interactions, a sufficient amount of experimental data, an exceptional performance of the detector hardware and reconstruction algorithms, and a powerful analysis approach to extract the $H \rightarrow \mu\mu$ signal. By the end of the LHC Run 2, the progress of the CMS experiment in these aspects had reached a point at which the sensitivity of the $H \rightarrow \mu\mu$ search was sufficient to witness an evidence of this decay for the first time [1].

The search performed using the full Run 2 dataset ($L = 137 \text{ fb}^{-1}$) features an elaborate analysis strategy that allowed to observe an excess of events over the background, consistent with the SM prediction for the $H \rightarrow \mu\mu$ decay, with an observed (expected) signal significance¹ of 3.0 (2.5) standard deviations. This result had become the first precise measurement of a Higgs boson coupling to the second generation fermions, opening a new chapter in Higgs physics. In the following sections, the analysis of the Run 2 $H \rightarrow \mu\mu$ search by CMS is described in detail.

¹↑ Expected significance denotes the degree of disagreement between the null (background-only) and the signal-plus-background hypotheses, while observed significance quantifies the disagreement of the observed data with the null hypothesis. More details on the statistical methods used in the $H \rightarrow \mu\mu$ analysis are given in Sections 4.6.3, 4.7.4, and 4.10.

4.1 Analysis strategy

A novel approach to the Run 2 $H \rightarrow \mu\mu$ search by the CMS collaboration, as compared to the earlier searches, was to split the data into mutually exclusive categories targeting different Higgs production modes — gluon-gluon fusion (ggH), vector boson fusion (VBF), production in association with a pair of top quarks ($t\bar{t}H$), and production in association with a vector boson (WH and ZH). A separate analysis strategy was developed for each category, based on two different approaches to background modeling, as described below.

Parametric background fit approach.

In the parametric fit approach, used in the ggH, $t\bar{t}H$, WH, and ZH categories, both signal and background are modeled with functions describing their invariant mass spectra. The background model is obtained from a fit of the observed data in the control region (mass sidebands²), while the signal model is obtained by fitting the mass spectrum of the $H \rightarrow \mu\mu$ signal MC. The benefit of this approach is that the background model is not subject to any systematic uncertainties associated with the MC simulation. On the other hand, the “true” parameterization of the background is unknown, and the limited number of events in the sidebands allows to describe the background with a variety of functions resulting in a comparable fit quality in the control region, but with slightly different background predictions in the signal region³. These differences may considerably bias the sensitivity to the signal, especially when the expected signal rate is low, as it is for the $H \rightarrow \mu\mu$ decay.

The sensitivity of the parametric fit approach is further improved by splitting the data into subcategories with different signal-to-background yield ratios and different widths of the signal mass peak. In the Run 2 analysis, this categorization was performed using multivariate analysis (MVA) techniques, such as boosted decision trees (BDT) and deep neural networks (DNN). Since in each subcategory the assumption of the smoothness of the background mass spectrum must be respected, it is crucial to ensure that the MVA models are completely uncorrelated with the dimuon mass. Otherwise, the signal-to-background ratio could be

²↑Mass sidebands are defined as $m_{\mu\mu} \in [110, 120) \cup (130, 150]$ GeV.

³↑Signal region is defined as $m_{\mu\mu} \in [120, 130]$ GeV.

artificially improved by selecting data specifically in the signal mass range. However, in that case a parametric background fit in the sideband region would no longer be possible.

The details of the parametric fit approach specific in each of the analysis categories are further discussed in Sections 4.7–4.9.

MC template-based approach.

An alternative approach is used in the VBF category. Here both the signal and the background models are estimated using the prediction of the MC simulation. In order to achieve a high level of sensitivity, a one-dimensional observable is constructed using a DNN, which takes as input the kinematics properties of the decay products, as well as other event properties that have discriminating power. The events in both data and MC are split into multiple non-overlapping bins based on the DNN score, such that the signal-to-background ratio is different in different bins. The resulting histograms are referred to as *templates*. The signal significance and other statistical parameters are then obtained by fitting the MC templates to the templates obtained from the observed data, using the signal normalization as a floating parameter, and taking into account statistical and systematic uncertainties. The fit procedure is described in more detail Section 4.6.3.

Since the smoothness of the background mass spectrum is not assumed in this approach, the DNN models are not required to be uncorrelated with the dimuon mass. On the contrary, the dimuon mass is used as one of the inputs for the DNN, since it provides the highest level of signal-background discrimination among the event properties.

The MC template-based method is advantageous in the regions of the phase space where the signal-to-background ratio is high, and a considerable sensitivity to the signal is achievable even in the presence of large uncertainties on the background MC. The main sources of the systematic bias in this approach arise from inaccuracies in the MC modeling; therefore, ensuring a high level of agreement between data and MC, as well as a good understanding of the discrepancies, is of a particular importance.

The details of the MC template-based strategy in the analysis category targeting the VBF production mode are discussed in Section 4.6.

4.2 Datasets

The full Run 2 dataset used in the $H \rightarrow \mu\mu$ analysis corresponds to an integrated luminosity of $L = 137 \text{ fb}^{-1}$. The starting point is a **SingleMuon** dataset, a collection of events passing at least one of the software triggers that require a presence of at least one muon in the event⁴. The dataset combines three data-taking periods of the CMS experiment — 2016, 2017, and 2018 — which feature slightly different data-taking conditions, and thus must be analyzed separately.

Various MC datasets were produced to model all known SM processes that have a non-negligible contribution to the backgrounds in the analysis categories. In addition, MC datasets were produced for all signal processes of interest. To ensure an accurate modeling, MC samples were generated separately for each period, taking into consideration different calibration and alignment conditions of subdetectors, different pile-up conditions, and other aspects.

The MC simulation is performed in multiple steps. First, events are generated at the matrix element level using either the MADGRAPH5_aMC@NLO [112] or the POWHEG [113] generator. Then, the events are processed by either PYTHIA [114] or HERWIG [115] generators, which model fragmentation and hadronization of the partons in the simulated events. Finally, the generated events are processed through a full simulation of the CMS detector implemented using GEANT4 [116, 117], in order to obtain a realistic detector response in the same format as that from collision data. The trigger algorithms, object reconstruction and identification algorithms, and other processing sequences are then applied in exactly the same way to both collision data and MC simulation events. The full list of the MC datasets used in the analysis can be found in Appendix A.

⁴↑It is certainly possible to select the events for the $H \rightarrow \mu\mu$ analysis using dimuon triggers. However, the offline algorithms reconstruct muons with a higher efficiency, as compared to the online reconstruction in the trigger algorithms. Therefore, the $H \rightarrow \mu\mu$ event selection efficiency is higher if only one of the muons is required to pass the single muon trigger, while the other one is thoroughly reconstructed by the offline algorithms.

4.3 Event selection and categorization

The events pre-selected for all categories of the analysis are required to contain a pair of oppositely charged muons within the pseudorapidity coverage of the CMS muon system ($|\eta| < 2.4$), and with transverse momenta exceeding 20 GeV. The transverse momentum of the leading muon is additionally required to exceed 26 GeV (29 GeV for the 2017 data-taking period) in order to reach the efficiency plateau of the HLT_IsoMu24⁵ trigger (HLT_IsoMu27 for 2017). Both muons are required to be isolated in order to reject non-prompt or misidentified muons, by requiring the summed momenta of particles located at a close angular distance from a muon to not exceed 25% of the muon momentum in the transverse direction.

The hadron jets considered in the analysis are reconstructed using the anti- k_T algorithm [101] with a distance parameter of $R = 0.4$; they are required to be within the pseudorapidity acceptance range of the HCAL and HF ($|\eta| < 4.7$), and only the jets with the transverse momentum exceeding 25 GeV are considered. Situations in which one of the selected muons is considered a part of a selected jet are already largely avoided by the muon isolation requirements, however, it is additionally enforced that all selected jets are at an angular distance of at least $\Delta R = 0.4$ from each of the selected muons.

The categorization of the selected events is based on the topology and kinematics of the final state, as illustrated in Figure 4.2. The events in the $t\bar{t}H$ category are required to have either one b -tagged jet passing the “medium” working point of the DeepCSV tagger [118], or two jets passing the “loose” working point. The VH category includes events without b -tagged jets, but with at least one lepton (muon or electron) in addition to the main muon pair. Out of the events not included in the $t\bar{t}H$ and VH categories, those containing a pair of jets passing the selection criteria mentioned above with invariant mass of at least 400 GeV, pseudorapidity separation $\Delta\eta_{jj} > 2.5$, and the leading jet p_T exceeding 35 GeV, are classified into the VBF category. The rest of the events are included in the ggH category.

Such a categorization allows for a considerable share of the $H \rightarrow \mu\mu$ events to be correctly classified by their production modes. However, due to the limited knowledge of the nature of jets in an event, there is a possibility for an overlap between different categories (for

⁵↑ HLT_IsoMu24 trigger requires an event to have at least one isolated muon with $p_T > 24$ GeV.

example, when a pile-up jet is mistaken for a jet produced in the VBF process). Thus, each analysis category is also expected to include a certain amount of the $H \rightarrow \mu\mu$ events from the production channels that are not specifically targeted by that category.

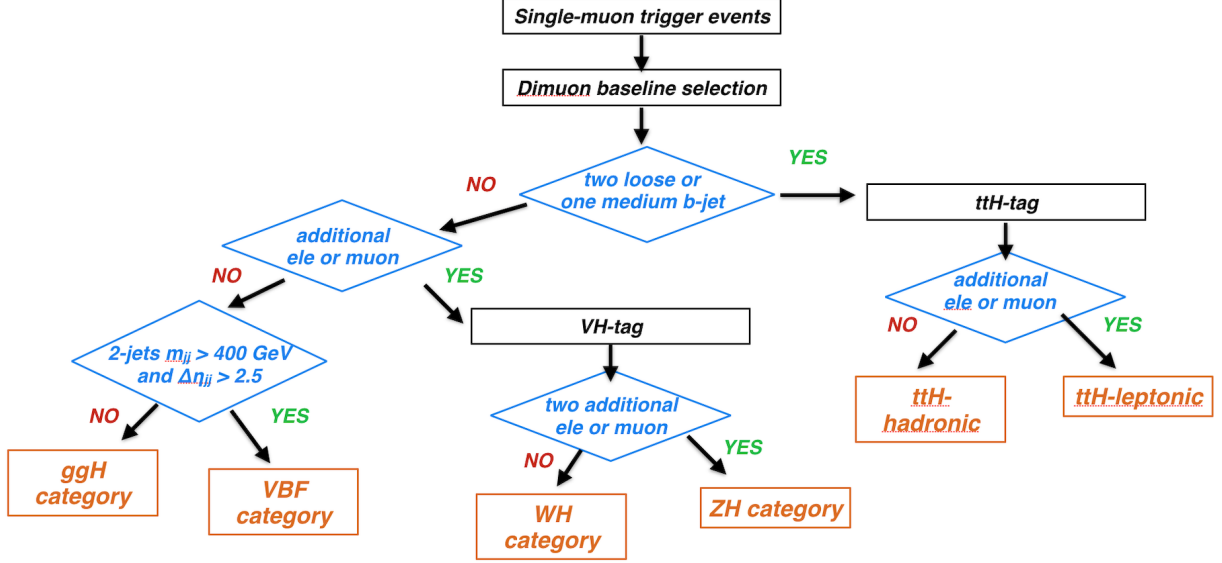


Figure 4.2. Event categorization in the Run 2 $H \rightarrow \mu\mu$ analysis [1].

4.4 Corrections to data and MC

While the simulated MC datasets provide an adequate description of the SM processes, there are certain differences between the distributions of the observed data and simulation, which may bias the results of the analysis. Therefore, it is essential to apply corrections to some of the observables in data and MC, in order to suppress various biases in detector response, mismodeling of physical observables in simulation, and to ensure that the simulation correctly models the key features of the signal and background processes. Some of the most important corrections applied in the $H \rightarrow \mu\mu$ analysis are described below.

- **Muon momentum scale and resolution correction (Rochester correction).**

The scale and resolution of the reconstructed muon momentum differs between the MC simulation and the observed data. The imperfections in muon reconstruction translate into a shift in the scale of the dimuon mass resonances, as well as a difference in the dimuon mass resolution between data and MC. The Rochester correction [119]

is derived from the assumption that the dimuon invariant mass peak corresponding to the Z boson in the Drell–Yan process should be aligned with its PDG value of 91.19 GeV [120] for both data and MC. Additionally, the difference in the dimuon mass resolution between data and MC is mitigated by smearing the muon momenta in the MC.

- **Final state radiation (FSR) recovery.** In approximately 8% of the $H \rightarrow \mu\mu$ events, at least one of the muons radiates a photon in the final state. Because of this, the reconstructed momentum of the muon decreases, and the dimuon invariant mass is shifted towards the low-mass region of the spectrum. The effect of the FSR radiation is corrected by recovering the original 4-momentum of the muon before the photon emittance. This is done by searching for an isolated photon compatible with the expected FSR kinematics within a close angular distance from the muon, and adding the 4-momenta of the two particles. The procedure is done before applying the muon isolation requirement, in order to prevent a situation in which a muon fails the isolation requirement solely due to the FSR photon emittance. The updated isolation requirements lead to $\sim 2\%$ increase in the signal yield, while the overall improvement to the dimuon mass resolution due to FSR recovery is around 0.5%.
- **GeoFit correction.** One of the reasons for the deteriorated muon p_T resolution is the fact that the muon tracks are produced by fitting the detector hits from the tracker and the muon system, but without an explicit constraint on the vertex from which the muon originates. In the $H \rightarrow \mu\mu$ events, due to the short lifetime of the Higgs boson, both muons are expected to originate from the collision point. Thus, the distance from the point of closest approach of the muon track to the beamspot is expected to be correlated with the muon momentum resolution: the farther the muon track is displaced from the beamspot, the more its reconstructed momentum is expected to deviate from its true value. The GeoFit correction [121] aims to improve the muon p_T resolution by correcting the muon p_T in a way to decrease its correlation with the distance to the beamspot in the transverse plane. Effectively, this approach is similar to including the beamspot location into the muon trajectory fit, but the computation

of the GeoFit correction is much faster, and it does not require re-fitting the trajectory. The application of the GeoFit correction improves the muon p_T resolution by 8–15%, and the dimuon mass resolution by 4–11%, depending on the muon pseudorapidity [2].

- **Lepton scale factors.** Another method to improve the agreement between data and MC is derived using the efficiencies of the muon identification and isolation requirements, as well as the efficiency for a given muon to pass a single muon trigger. These efficiencies, measured on both data and MC, are used to compute the corresponding MC/Data scale factors in p_T and η bins, which are then applied as weights to the MC events, improving the agreement between data and MC.
- **b -tag weights.** The probability distribution for a jet to come from the hadronization of a bottom quark is derived in CMS using a supervised machine learning algorithm [118] trained on MC simulation. The b -tagging efficiency is calculated for data and MC in bins of jet p_T , η , in event samples enriched either in jets from b -quarks or in “light flavour” jets (light quarks and gluons). The b -tag weights are then derived using the bin-by-bin ratios of these efficiencies, and applied to MC in order to reduce the disagreement between data and MC.
- **Pile-up weights.** The MC simulated event samples are produced with a pile-up profile that is approximately similar to that measured in data, but does not exactly match it. A reweighting procedure is applied in order to correct for the remaining differences, using the distributions of the number of reconstructed vertices in data and MC as a reference.
- **Other corrections.** Other corrections used in the $H \rightarrow \mu\mu$ analysis include the corrections to jet energy scale and resolution, weights used to correct the data and MC efficiencies of the “pile-up ID” and “quark-gluon likelihood” jet classifiers, as well as various corrections developed to account for the limitations of the $H \rightarrow \mu\mu$ signal simulation.

4.5 Systematic uncertainties

The modeling of the signal and background processes used in all analysis categories is subject to systematic uncertainties, which translate into the final uncertainties on the results of the $H \rightarrow \mu\mu$ search. Most of these uncertainties are modeled as nuisance parameters with log-normal or Gaussian external constraints. Some of the leading sources of systematic uncertainty are described in this section.

4.5.1 Theoretical uncertainties

The Higgs boson production cross section for the various production modes, and their corresponding uncertainties, are taken from Ref. [43]. These include uncertainties in the choice of the parton distribution function (PDF), uncertainties on the QCD renormalization and factorization scales, and the uncertainty in the prediction of the branching ratio $\mathcal{B}(H \rightarrow \mu\mu)$. For the ggH process, seven independent additional sources are included to account for the uncertainty in the modeling of the $p_T^{\mu\mu}$ distribution, the number of jets in the event, and its contamination in the VBF selected region, as described in Ref. [43]. The magnitude of these uncertainties for ggH events in the VBF category varies from about 15% to 25%, while in the ggH category the total uncertainty in the ggH yield ranges from 6% to 12% depending on the subcategory. Similarly, for the VBF process, uncertainties in the modeling of the $p_T^{\mu\mu}$, $p_T^{\mu\mu jj}$, jet multiplicity, and m_{jj} distributions are considered, accounting for a total uncertainty on the VBF signal prediction of about 2–4% in both ggH and VBF categories.

The leading source of uncertainty in the VBF category arises from the limitations of the parton shower (PS) modeling. This uncertainty is assessed by varying several parameters that control the properties of the jets produced by the PYTHIA generator. A conservative PS uncertainty is assigned to the Zjj -EW background and VBF signal, defined as the absolute difference between PYTHIA (dipole shower) and HERWIG (angular-ordered shower) predictions in each bin of the DNN output score in that category.

4.5.2 Experimental uncertainties

Several sources of experimental uncertainty are taken into account for both signal and background processes. These include the uncertainty in the measurement of the integrated luminosity, in the modeling of the pile-up conditions during data taking, in the measurement of the muon selection and trigger efficiencies, in the muon momentum scale and resolution, in the efficiency of vetoing b -quark jets, and in the jet energy scale and resolution.

The integrated luminosities of the 2016, 2017, and 2018 data-taking periods are individually known with uncertainties in the 2.3–2.5% range [122–124], while the total integrated luminosity has an uncertainty of 1.8%. The effects of other uncertainties are measured separately in all event subcategories in the ggH , $t\bar{t}H$ and VH channels, as well as in the VBF category, as listed in Table 4.1.

Table 4.1. Experimental uncertainties in the $H \rightarrow \mu\mu$ event (sub-)categories.

Source of uncertainty	Uncertainty per (sub-)category
Muon and electron selection efficiencies	0.5–3%
Jet energy scale and resolution	0.5–6%
Muon momentum scale and resolution	0.1–0.8%
modeling of the pile-up conditions	0.2–2%
Efficiency of identifying and vetoing b -quark jets	0.1–3%

In the category targeting the VBF production mode, there is an additional source of uncertainty arising from the limited size of the simulated samples. It is taken into account by allowing each bin of the total background template to vary within the corresponding statistical uncertainty using the Barlow–Beeston lite technique [125]. These uncertainties are uncorrelated across the bins of the DNN templates used in the fit.

In the categories where the data-driven method of background estimation is used, a systematic uncertainty associated with the choice of the parametric background fit function is considered. A discrete profiling method [126] is employed to estimate this uncertainty, treating the choice of the fit function as a discrete nuisance parameter in the signal extraction.

4.6 VBF category

The analysis in the VBF category of the Run 2 $H \rightarrow \mu\mu$ search was implemented independently by multiple research groups, in order to ensure the robustness of the approach and avoid errors in the final results. One of the implementations, developed by the author of this thesis, is described in this section. It has been verified that this implementation is consistent with the implementation featured in Ref. [1] to a high degree, and the remaining differences are understood. The comparison between the results obtained with the two implementations is given at the end of the section.

4.6.1 Event selection and Data/MC agreement

The phase space of the category targeting the VBF production channel is defined by the pre-selection of the muon pair, mentioned in Section 4.3, a requirement for the event to not contain additional leptons or b -jets, and a requirement for the presence of a jet pair that passes the VBF kinematic selections ($m_{jj} > 400$ GeV, $\Delta\eta_{jj} > 2.5$, $p_T^{j1} > 35$ GeV). Such a selection allows to suppress a substantial share of the Drell–Yan and $t\bar{t}$ backgrounds, however, the amount of the remaining backgrounds is still significant. The main backgrounds in the phase space of the VBF category are:

- Drell–Yan (DY) events in which the jets are either produced via initial state radiation of gluons, or picked up from pile-up, or misreconstructed from the detector noise;
- Events with electroweak production of a Z boson through a VBF process with a topology similar to that of the VBF $H \rightarrow \mu\mu$ process (denoted as Zjj -EW or EWK in the following text and plots).

The analysis sensitivity is enhanced by considering the differences in the distributions of event properties between the VBF $H \rightarrow \mu\mu$ signal and the backgrounds. Although each individual distribution does not have enough discriminating power to achieve a high signal significance, a combination of multiple properties can provide a higher sensitivity than each of its constituents. At the same time, the amount of observed data is not sufficient to perform a fit in multiple dimensions, therefore the combined sensitivity must be extracted

from a one-dimensional observable. Such an observable is constructed using a deep neural network (DNN), which takes event properties as inputs, and is trained using signal and background MC to produce a one-dimensional output score which maximizes the signal-background discrimination. The use of the DNN score as a sensitive observable is motivated by the fact that it approximates the likelihood ratio distribution, which is the most sensitive test statistics for hypothesis testing (see Section [B.1.1](#) for a more detailed discussion).

Before training the DNN model, it is important to ensure that the distributions of input features in the background MC match the corresponding distributions for the observed data in the phase space region of interest; otherwise, the observable constructed using the DNN may be subject to a systematic bias. The comparison of data and MC distribution for some of the kinematic parameters of the muon pair and the jet pair are shown in Figures [4.3](#) – [4.5](#). While the agreement between MC and observed data is improved by applying the corrections described in Section [4.4](#), certain discrepancies remain in the distributions of some of the variables. Most of the discrepancies were tracked down to the imperfections in modeling and reconstruction of the hadron jets. Such imperfections affect not only the distributions of the jet-related variables, but also the kinematic distributions of the muons, due to the fact that all events selected for the VBF category are required to pass the selections defined using the jet pair kinematics. The discrepancies in the distributions are taken into account by considering the variations of the jet kinematic properties, which arise from the uncertainties on the jet energy scale and resolution in the MC, and propagating them to the variations of the DNN output score distributions.

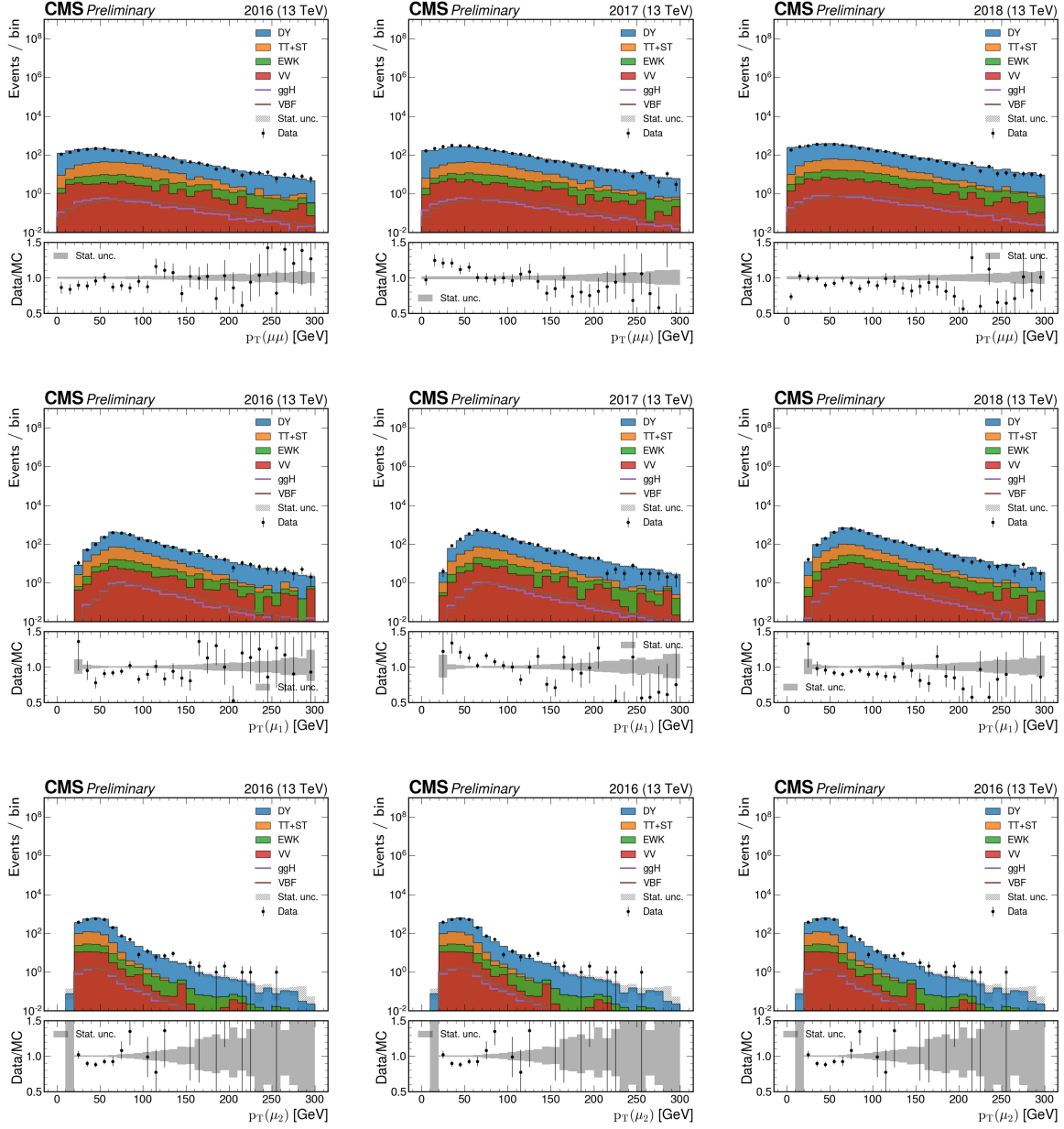


Figure 4.3. Distributions of dimuon p_T (top), leading muon p_T (middle) and subleading muon p_T (bottom) for 2016 (left), 2017 (center) and 2018 (right).

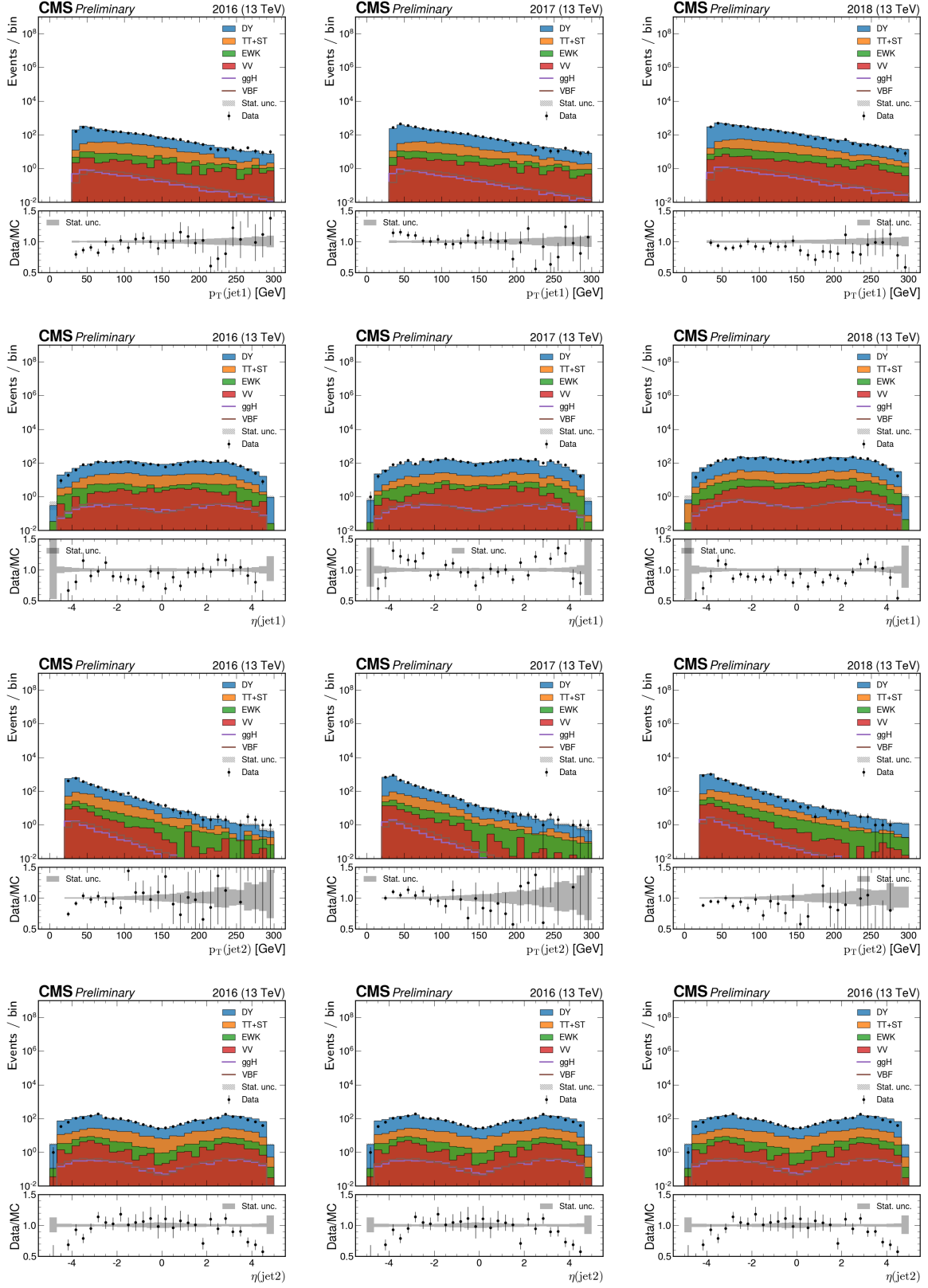


Figure 4.4. Distributions of p_T and η for the leading jet (top two rows), and for the subleading jet (bottom two rows) and subleading muon p_T (bottom) for 2016 (left), 2017 (center) and 2018 (right).

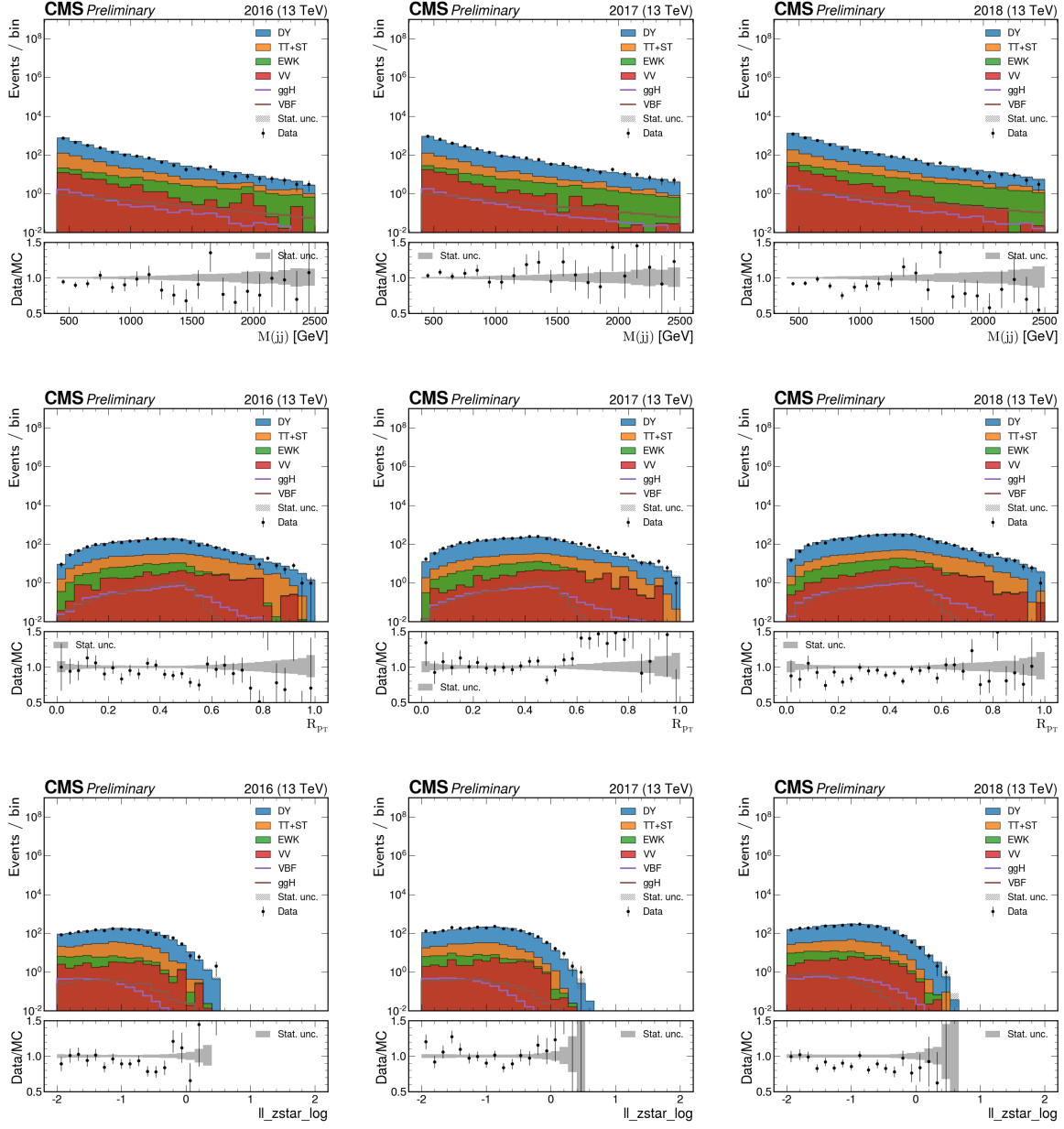


Figure 4.5. Distributions of dijet invariant mass (top), p_T -balance ratio (R_{p_T}) (middle) and logarithm of the Zeppenfeld variable z^* (bottom) for 2016 (left), 2017 (center) and 2018 (right).

4.6.2 DNN classifier

The input variables for the DNN are defined in a way to describe an event in the VBF category as comprehensively as possible. In addition to the kinematics of the dimuon and dijet pairs, their combined properties are considered, such as the p_T -balance ratio R_{p_T} that quantifies the relation between transverse momenta of the jets and the dimuon pair, and the Zeppenfeld variable z^* [127], which quantifies whether jets are located within the pseudorapidity gap between the two muons. For each of the two jets, aside from their kinematics, the value of the quark-gluon likelihood discriminator (QGL) [128] is considered, which allows to extract a certain degree of sensitivity using jet substructure. Finally, the sensitivity is enhanced by considering the soft jet activity in an event. In order to achieve a more stable training, the datasets corresponding to the different data-taking periods are combined, and the “year” variable denoting these periods is added as an input in order to take advantage of the subtle differences due to the changing run conditions. The full list of features is given below, and the relative importance of these features is discussed in Section 5.2.1:

- Dimuon system kinematics: $m_{\mu\mu}$, p_T , $\log p_T$, η
- Angles in the Collins-Soper frame [129]: $\cos\theta_{CS}$, ϕ_{CS}
- Absolute and relative per-event dimuon mass resolution
- For each of the two selected jets: p_T , η , ϕ , QGL
- Dijet system kinematics: m_{jj} , $\log m_{jj}$, $\Delta\eta_{jj}$
- p_T -balance ratio: $R_{p_T} = \frac{|\vec{p}_T^{jj} + \vec{p}_T^{\mu\mu}|}{|\vec{p}_T^{j1}| + |\vec{p}_T^{j2}| + |\vec{p}_T^{\mu\mu}|}$
- Zeppenfeld variable [127]: $z^* = \frac{y^*}{|y_{j1} - y_{j2}|}$, where $y^* = y_{\mu\mu} - \frac{y_{j1} + y_{j2}}{2}$
- Angular separation between the dimuon pair and the closest jet: $\min\Delta\eta_{\mu\mu,j}$
- Soft activity variables: $N_{\text{jets}}^{\text{soft}}(p_T > 5 \text{ GeV})$, $H_T^{\text{soft}}(p_T > 2 \text{ GeV})$
- Year (2016/2017/2018)

A variety of different DNN architectures were considered in different implementations of the VBF $H \rightarrow \mu\mu$ analysis. The DNN architecture used to produce the final results in Ref. [1] contains five sub-networks with different input features and training datasets, while the network used in the implementation illustrated in this chapter features a simpler, three-layer architecture. The details of the DNN architecture and the training settings are given in Section 5.2.1. The discriminating power of the DNN is illustrated in Figure 4.6.

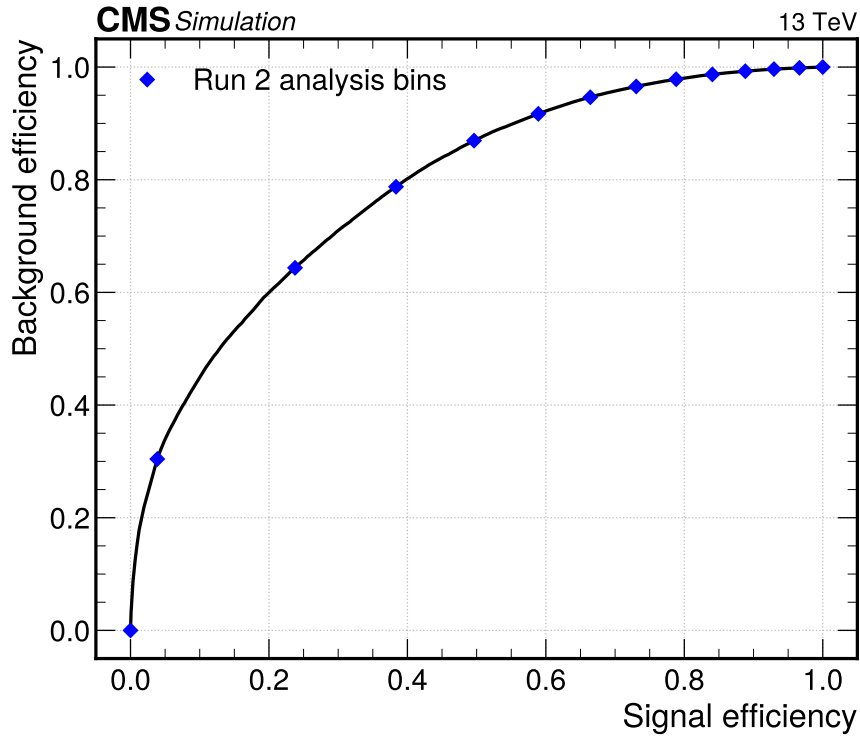


Figure 4.6. Receiver operating characteristic (ROC) curve for the DNN used in the VBF category. Signal (background) efficiency is defined as signal (background) yield below a given DNN score cutoff, divided by the total signal (background) yield. The point (0,0) corresponds to the lowest DNN score, while (1,1) corresponds to the highest score. Blue markers correspond to the bin boundaries used in the Run 2 analysis.

4.6.3 Fit procedure

The trained DNN is evaluated on signal and background MC datasets, as well as on the observed data. Two regions are defined based on the dimuon invariant mass (note that the definition is different from that in other categories):

- Sidebands (VBF-SB): $m_{\mu\mu} \in [110, 115) \cup (135, 150]$ GeV.
- Signal region (VBF-SR): $m_{\mu\mu} \in [115, 135]$ GeV.

Events belonging to the VBF-SR region are divided into non-overlapping bins based on the DNN score, independently for each data-taking period. These bins are defined to achieve optimal sensitivity, while minimizing the total number of bins. The total number of bins for each data-taking period is fixed to 13, the highest-score bin is required to contain an expected VBF $H \rightarrow \mu\mu$ yield of 0.6, and the yields in other bins are required to scale linearly with the bin number. During the evaluation of the DNN in the VBF-SB region, the value of the dimuon invariant mass is set to 125 GeV for all events. Since the dimuon mass is largely uncorrelated with other DNN inputs, this method allows to achieve an output distribution similar to that in the VBF-SR region, which makes it possible to use the same binning scheme as the one defined for the VBF-SR region.

The signal and background yields in the MC templates are subject to systematic uncertainties, as described in Section 4.5. These uncertainties are quantified with *nuisance parameters*, which allow for variations of the MC yields either by the same amount in all bins (*rate variations*), or by a different amount in each individual bin (*shape variations*). In order to constrain the values of the nuisance parameters, the background MC templates are fit to data in the VBF-SB region, where the signal contribution is negligible. The fit is performed using a *profile likelihood* approach [130], which allows to obtain the optimal values of nuisance parameters. These values are then used to constrain the uncertainties in the VBF-SR regions, since most of the systematic uncertainties are assumed to be fully correlated between VBF-SR and VBF-SB. The MC templates before and after this procedure are referred to as *pre-fit* and *post-fit*, respectively.

The distributions of the binned DNN score before and after the fit are shown separately for each data-taking period in Figures 4.7–4.9. In all cases, a $\tanh^{-1}$ transformation is applied to the DNN output scores, in order to show the features of the distributions. This transformation has no effect on the results of the statistical analysis, as long as the bin boundaries are adjusted accordingly. The post-fit distributions for the combination of all three data-taking periods are shown in Figure 4.10.

Another fit is performed using the updated templates in the VBF-SR region to extract the signal. In this case, the background-only and signal-plus-background hypotheses are fit to the observed data, with signal strength as a free parameter. The signal strength denotes the ratio of the observed signal rate to the rate predicted by the SM (see definition in Eq. 4.6). Because of significant variations in the detector response to forward jets during different data-taking periods, the fit is performed separately for data collected in 2016, 2017, and 2018, but the signal strength parameter is shared across all three periods.

A significant fraction (about 30–35%) of the DY background populating bins with low DNN score is comprised of events in which either the leading or the subleading jet is in the forward region of the detector ($|\eta| > 3.0$) and not matched to a generator-level jet. These jets originate either from the soft emissions produced by the parton shower or from pile-up interactions. The normalization of this component of the Drell–Yan background is assigned a flat prior and left floating during the fit, which allows it to be directly constrained by the observed data events with low DNN score. The normalization of the remaining DY component with at least two matched jets is taken from the simulation and constrained similarly to the other background processes.

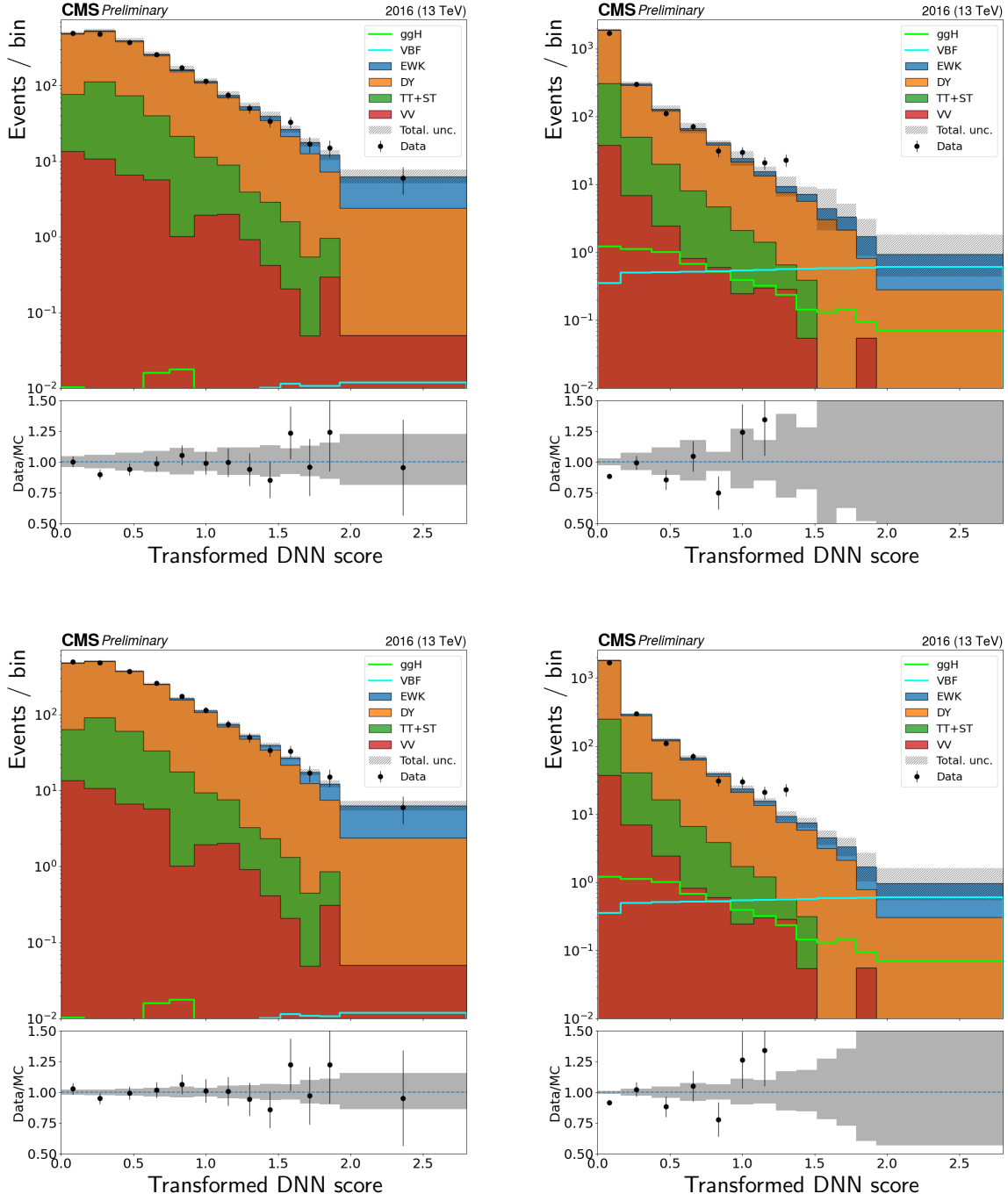


Figure 4.7. Pre-fit (top row) and post-fit (bottom row) DNN score templates in VBF-SB region (left) and VBF-SR region (right) shown for the 2016 data-taking period.

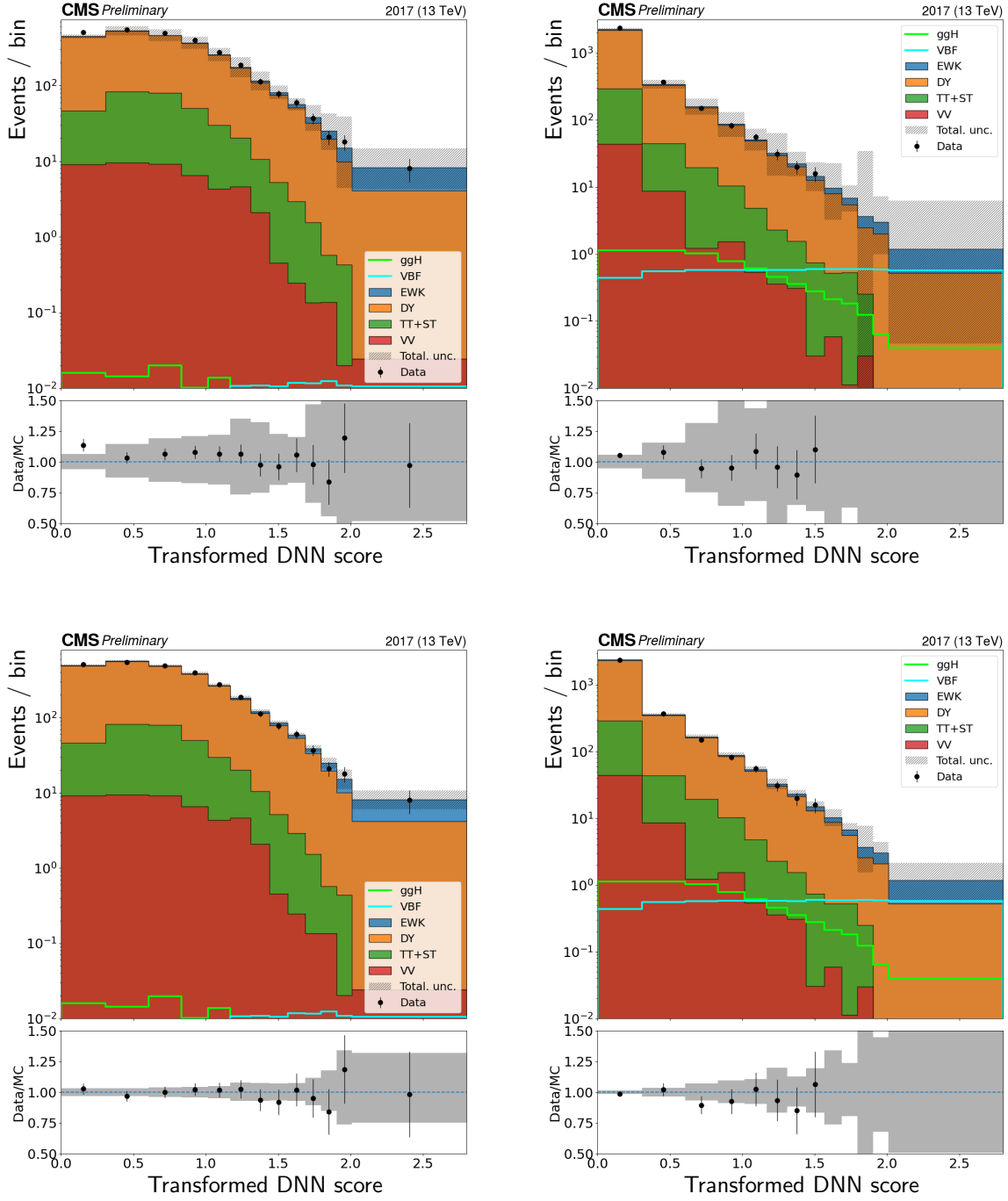


Figure 4.8. Pre-fit (top row) and post-fit (bottom row) DNN score templates in VBF-SB region (left) and VBF-SR region (right) shown for the 2017 data-taking period.

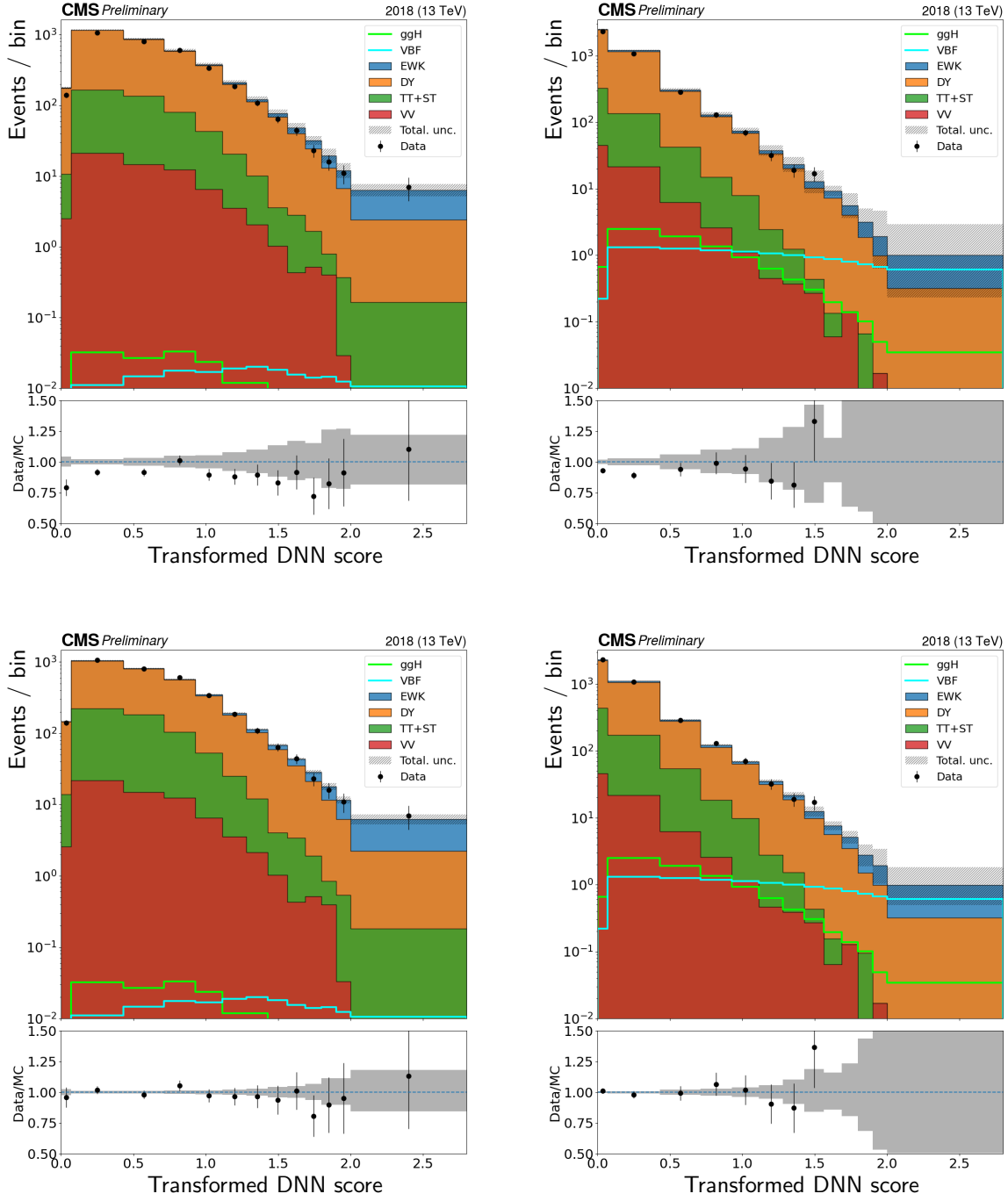


Figure 4.9. Pre-fit (top row) and post-fit (bottom row) DNN score templates in VBF-SB region (left) and VBF-SR region (right) shown for the 2018 data-taking period.

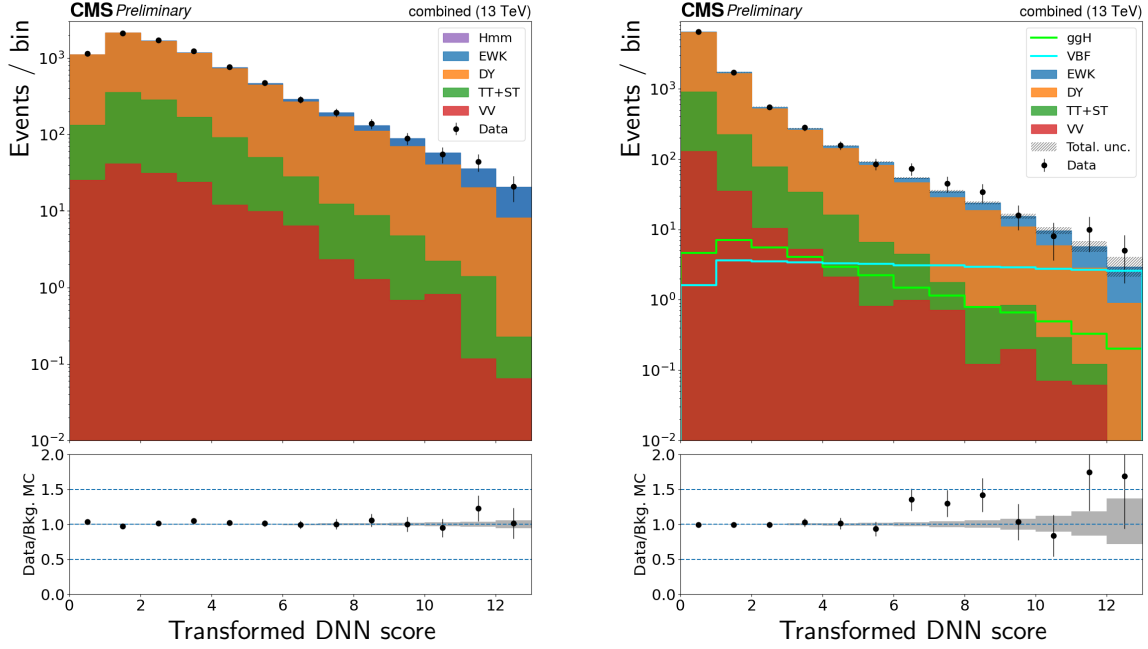


Figure 4.10. Post-fit DNN score templates in VBF-SB region (left) and VBF-SR region (right), combined for all three data-taking periods.

Figure 4.11 illustrates the effect of the main systematic uncertainties on the signal strength after the nuisance parameters are constrained. The highest impacts correspond to the selection of the generator settings to model the parton shower uncertainties for the VBF signal and Zjj -EW background, as well as the normalization of the Drell–Yan background and the uncertainties related to jet energy scale and resolution.

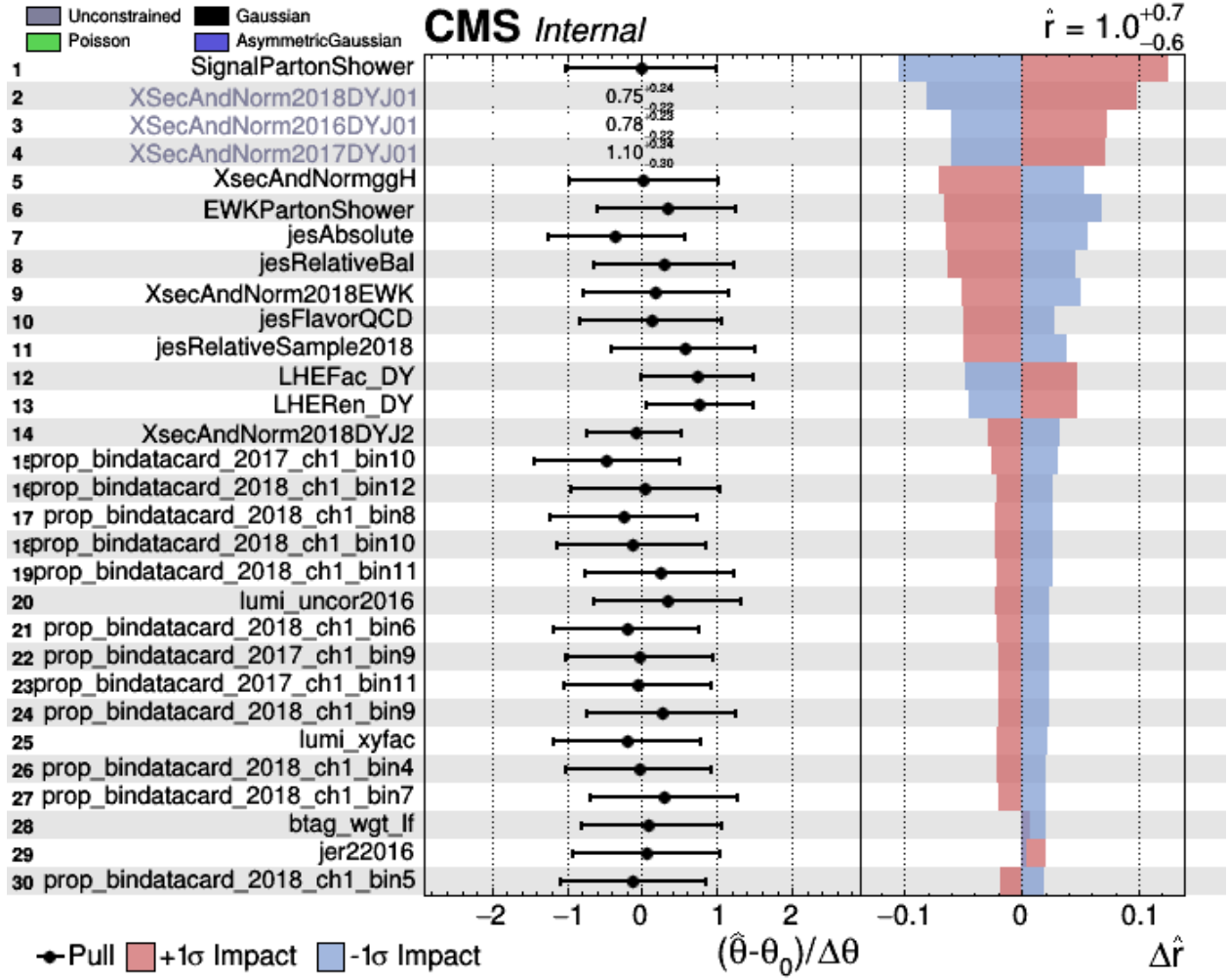


Figure 4.11. Impacts of the most important systematic uncertainties on the signal strength fit in the VBF category.

4.6.4 Comparison between implementations

The implementation of the VBF category of the $H \rightarrow \mu\mu$ analysis illustrated in this section is synchronized with the implementation featured in Ref. [1] to a high level of precision. The event yields for all MC processes and observed data in all data-taking periods are consistent in the two implementations, with differences not exceeding 1%. The DNN models and training settings are different between the setups, but the binning of the output score is defined such that the signal yields in all DNN bins are exactly the same in both implementations. This ensures that the differences in the final results are caused only by the choice of the DNN

model, but not by any other steps of the analysis. The implementation of the systematic uncertainties is also synchronized.

During the synchronization process, the observed data yields in the high-score DNN bins were not considered, in order to avoid the possibility of biasing the observed results. Thus, the figure of merit for the quality of synchronization was the expected post-fit significance. The comparison of the results obtained with the two implementations for the $m_H = 125.0$ GeV and $m_H = 125.38$ GeV are quoted in Table 4.2.

Table 4.2. Expected and observed significance for two implementations of the VBF $H \rightarrow \mu\mu$ analysis.

		Ref. [1]	This thesis
$m_H = 125.0$ GeV	Expected (post-fit)	1.81σ	1.77σ
	Observed	1.72σ	1.92σ
$m_H = 125.38$ GeV	Expected (post-fit)	1.77σ	1.76σ
	Observed	2.40σ	2.50σ

The expected results at $m_H = 125.38$ GeV show negligible difference, while at $m_H = 125.0$ GeV the difference is more substantial. This difference was tracked down to the method of DNN model selection, described in more detail in Section B.1.6. After synchronizing the model selection method, the expected significance at $m_H = 125.0$ GeV was also synchronized between the implementations.

The difference in the observed results is caused primarily by the event migration between the bins of the DNN output score. This is illustrated in Figure 4.12, where the two DNN setups are evaluated on the same sets of the observed data. The two DNN models are highly correlated, and the migration of events in most cases occurs in 2–3 neighboring bins, which suggests that the models learn the same information during training.

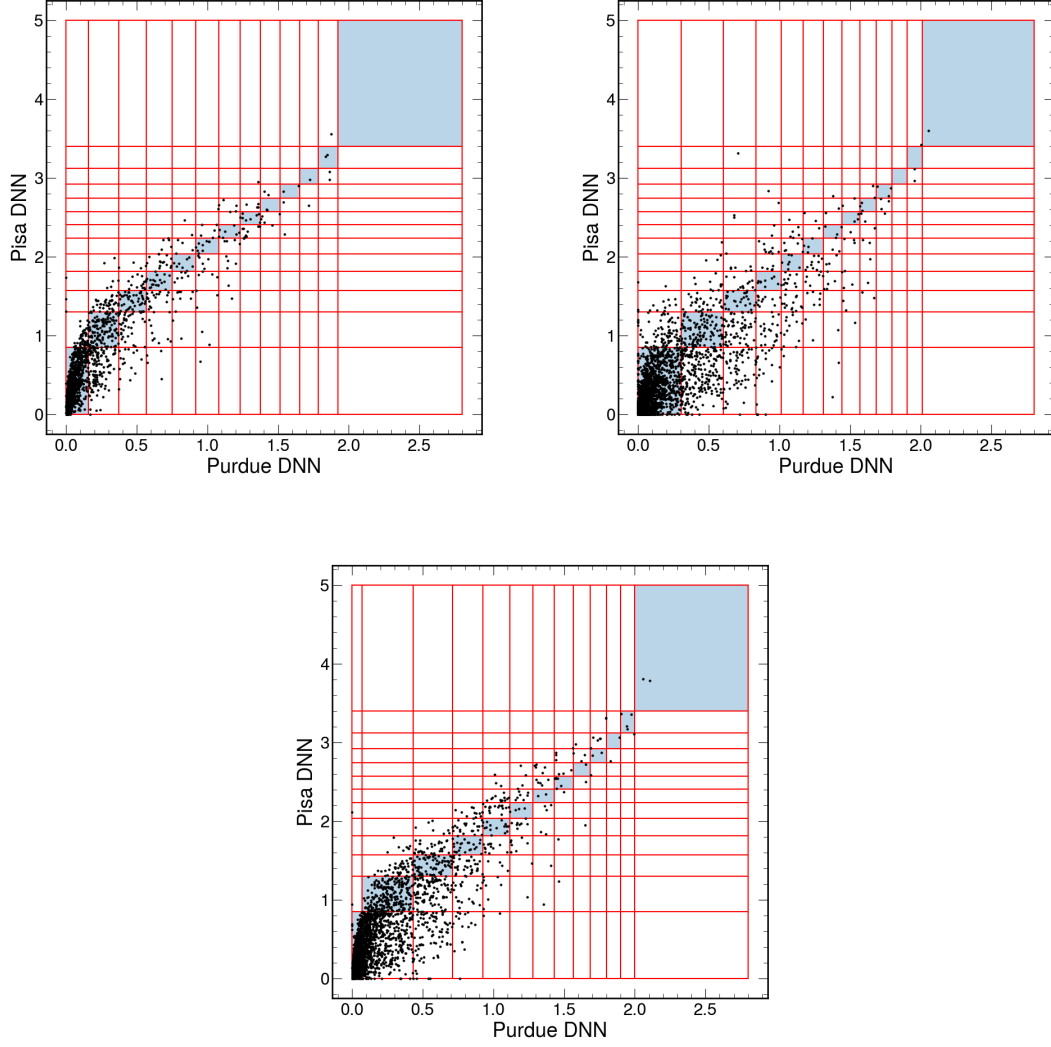


Figure 4.12. Correlation of the DNN models evaluated using 2016 (top left), 2017 (top right), 2018 (bottom) datasets. The x-axis corresponds to the DNN score in the implementation illustrated in this thesis, while the y-axis shows the score distribution for the model used in Ref. [1]. The red lines are the edges of the DNN bins, while the blue cells denote the DNN bins that correspond to each other between the two implementations (the signal yield in these bins is synchronized by design).

4.7 ggH category

The category targeting the ggH production mode includes the pre-selected events that do not feature additional leptons or b -jets, and do not pass the VBF selection criteria. The analysis strategy in this category employs a data-driven method of background estimation, using the dimuon invariant mass as a sensitive observable. The background is dominated by the off-shell production of the Z boson in the Drell–Yan process, which follows a smooth, monotonically falling distribution. The signal, in turn, forms a peak centered at $m_{\mu\mu} = 125$ GeV in the dimuon mass spectrum, with a width of several GeV, defined by the detector resolution. Both signal and background models are constructed using parametric functions, fitting the mass sidebands in the observed data for the background models, and fitting the MC prediction for the signal models. The parameters of the signal model are fixed in the final fit to data, and only the total signal normalization is left floating.

4.7.1 BDT categorization

The sensitivity of the data-driven approach is enhanced by splitting the events into subcategories with different levels of sensitivity to the signal. For this purpose, a multivariate discriminant based on boosted decision trees (BDT) is trained using the MC simulation of both signal and background. To account for the evolution in the detector response during the data-taking periods, the BDT discriminant is trained separately for the 2016, 2017, and 2018 simulated samples using the TMVA package [131], resulting in three independent BDT outputs. The input variables are chosen such that the BDT discriminants are effectively uncorrelated with $m_{\mu\mu}$. As mentioned in Section 4.1, this is done in order to ensure that the background mass spectrum remains smooth in the subcategories defined using the BDT score.

The BDT discriminants include input variables that describe the production and decay of the dimuon system, namely $p_T^{\mu\mu}$, $y_{\mu\mu}$, ϕ_{CS} , and $\cos\theta_{CS}$. In addition, the η and $p_T/m_{\mu\mu}$ of each of the two muons are also included. In order to increase the signal-to-background separation for events in which the ggH signal is produced in association with jets, the BDT discriminants also take into account the p_T and η of the leading jet in the event with $p_T > 25$ GeV and

the absolute distance in η and ϕ between the jet and the muon pair. For events with two or more jets with $p_T > 25$ GeV in the final state, additional inputs are included: the m_{jj} , $\Delta\eta_{jj}$, and $\Delta\phi_{jj}$ of the two highest p_T jets. Furthermore, the Zeppenfeld variable defined in Section 4.6.2, and the angular separation $(\Delta\eta, \Delta\phi)$ between the dimuon system and each of the two leading jets are also included, which target residual VBF signal events in the ggH selected region. Lastly, the total number of jets in the event with $p_T > 25$ GeV and $|\eta| < 4.7$ is also used as input to the BDT.

Only events with $m_{\mu\mu}$ in the range 115–135 GeV are included in the training. Signal events in the BDT training are assigned a weight inversely proportional to the expected mass resolution, derived from the uncertainties in the p_T measurements of the individual muon tracks. This weighting improves the average signal $\sigma_{\mu\mu}/m_{\mu\mu}$ in the high-score BDT region by assigning increased importance to the high-resolution signal events, as shown in Figure 4.13 on the right.

Figure 4.13 (left) shows the BDT score distribution, comparing data to the prediction from simulation in events with $110 < m_{\mu\mu} < 150$ GeV, where the outputs of the individual BDTs obtained in each year are combined into a single distribution. The distributions for various signal processes (ggH, VBF, and $VH + t\bar{t}H$) are also shown. Five event subcategories are defined based on the output of these BDT discriminants. The subcategory boundaries are determined via an iterative process that aims to maximize the expected sensitivity of this analysis to $H \rightarrow \mu\mu$ decays of the SM Higgs boson. The optimized event categories are labelled as “ggH-cat1”, “ggH-cat2”, “ggH-cat3”, “ggH-cat4”, and “ggH-cat5” corresponding to signal efficiency quantiles of 0–30, 30–60, 60–80, 80–95, and >95%, respectively. The grey vertical bands in Figure 4.13 (left) indicate the small range of variation, among the data-taking years, of the BDT boundaries for the optimized event categories described above.

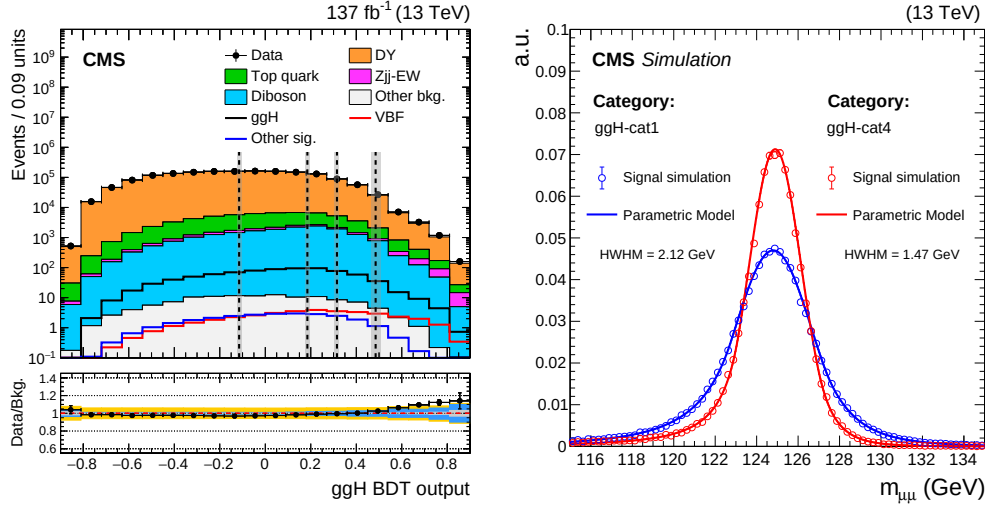


Figure 4.13. Left: distribution of the BDT score in the ggH category and the subcategory boundaries. Right: comparison of signal mass resolution in a low-score and a high-score BDT subcategories [1].

4.7.2 Signal modeling

The signal shape parameterization is required to correctly describe both the Gaussian-like peak of the distribution, and its elevated low-mass tail caused by the final state radiation (FSR). In the earlier iterations of the $H \rightarrow \mu\mu$ search, the shape was modeled either by a sum of two or three Gaussians [132], or by a sum of a Gaussian and a Crystal Ball function⁶ [134]. The main disadvantage of these approaches is the use of more than one Gaussian to describe the core of the distribution, as it could lead to unstable fits, interference between fit parameters, and a lack of physical sense behind using more than two parameters to describe the position and width of the peak. As an alternative, in the Run 2 $H \rightarrow \mu\mu$ analysis, a double-sided Crystal Ball function defined in Eq. 4.2 was used. Similarly to the regular Crystal Ball function, it uses a Gaussian core with power-law tails, but the parameters of the left and right tails are defined independently, allowing to correctly model the FSR tail without introducing an additional Gaussian.

⁶↑see Eq. 4.1 for the definition of the Crystal Ball function and Ref. [133] for the analysis in which it was introduced.

$$\text{CB}(x; \mu, \sigma, \alpha, n) = \begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & \frac{x-\mu}{\sigma} > \alpha \\ \left(\frac{n}{|\alpha|}\right)^n \cdot e^{-\frac{\alpha^2}{2}} \cdot \left(\frac{n}{|\alpha|} - |\alpha| - \frac{x-\mu}{\sigma}\right)^{-n} & \frac{x-\mu}{\sigma} \leq -\alpha \end{cases} \quad (4.1)$$

$$\text{DCB}(x; \mu, \sigma, \alpha_{L/R}, n_{L/R}) = \begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & -\alpha_L < \frac{x-\mu}{\sigma} < \alpha_R \\ \left(\frac{n_L}{|\alpha_L|}\right)^{n_L} \cdot e^{-\frac{\alpha_L^2}{2}} \cdot \left(\frac{n_L}{|\alpha_L|} - |\alpha_L| - \frac{x-\mu}{\sigma}\right)^{-n_L} & \frac{x-\mu}{\sigma} \leq -\alpha_L \\ \left(\frac{n_R}{|\alpha_R|}\right)^{n_R} \cdot e^{-\frac{\alpha_R^2}{2}} \cdot \left(\frac{n_R}{|\alpha_R|} - |\alpha_R| - \frac{x-\mu}{\sigma}\right)^{-n_R} & \frac{x-\mu}{\sigma} \geq \alpha_R \end{cases} \quad (4.2)$$

For a full comparison of different signal fit functions, the fits using various combinations of Gaussians, Crystal Ball and double-sided Crystal Ball functions are shown in Figure 4.14. Unlike other considered functions, the double-sided Crystal Ball function features only one Gaussian at its core, while at the same time providing an adequate fit quality.

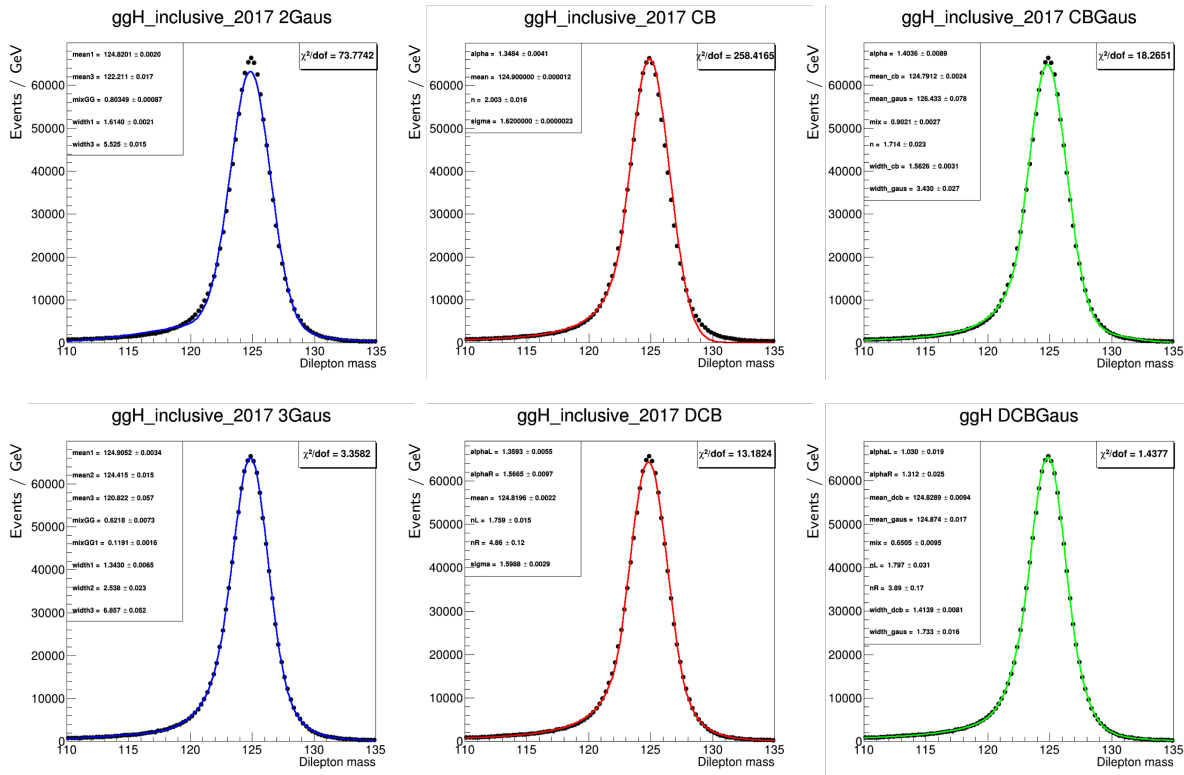


Figure 4.14. Different signal models considered to parameterize the mass spectrum of the ggH $H \rightarrow \mu\mu$ signal. Top row, from left to right: sum of two Gaussians, Crystal Ball, sum of Crystal Ball and a Gaussian. Bottom row, from left to right: sum of three Gaussians, double-sided Crystal Ball, sum of double-sided Crystal Ball and a Gaussian.

4.7.3 Background modeling

The background contribution in each subcategory is also modeled with parametric functions. No prior knowledge of the shape parameters of these functions or the yield of the total background is assumed. These parameters are therefore constrained directly by the observed data in the signal-plus-background fit. Since the background composition expected from the simulation is very similar across subcategories and largely dominated by the DY process, the background shape in $m_{\mu\mu}$ is similar in all event categories. There are, however, variations in the overall slope of the $m_{\mu\mu}$ spectrum across the BDT score categories. The function describing the background in each event category is therefore defined as the product of a “core” shape that is common among all event categories, with parameters correlated across categories, and a Chebyshev polynomial term (shape modifier) specific to each event category that modulates the core shape. This background modeling approach is referred to as the “core-pdf” method.

The core background shape is obtained from an envelope of three distinct functions, each containing three freely floating parameters:

- The modified form of the Breit–Wigner function (mBW)

$$\text{BWZRedux}(m_{\mu\mu}; m_Z, \Gamma_Z, a_1, a_2, a_3) = \frac{e^{a_2 \cdot m_{\mu\mu} + a_3 \cdot m_{\mu\mu}^2} \Gamma_Z}{(m_{\mu\mu} - m_Z)^{a_1} + (\Gamma_Z/2)^{a_1}} \quad (4.3)$$

where the parameters m_Z and Γ_Z are fixed to the measured Z boson mass of 91.19 GeV and width 2.49 GeV [23], and the parameters a_1 , a_2 , and a_3 are free to float.

- A sum of two exponentials.
- The product of a non-analytical shape derived from the FEWZ v3.1 generator [135] and a third-order Bernstein polynomial.

In a given subcategory, each of the three core functions is modulated by either a third-order (ggH-cat1 and ggH-cat2) or a second-order polynomial, with parameters uncorrelated across event categories.

The following strategy is adopted to estimate the uncertainty in the measured signal due to the choice of parametric function for the background model. In each event category, background-only fits to the data are performed using different types of functions: the mBW, a sum of two exponentials, a sum of two power-law functions, a Bernstein polynomial, the product between the nonanalytical shape described above and a Bernstein polynomial, the product between the “BWZ” function, defined as

$$\text{BWZ} \times \text{Bernstein}(m_{\mu\mu}) = \frac{e^{a \cdot m_{\mu\mu}} \Gamma_Z}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} \times \text{Bern}_n(m_{\mu\mu}) \quad (4.4)$$

and a Bernstein polynomial, and the “BWZ γ ” function defined as

$$\text{BWZ}\gamma(m_{\mu\mu}) = f \cdot \frac{e^{a \cdot m_{\mu\mu}} \Gamma_Z}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} + (1 - f) \cdot \frac{e^{a \cdot m_{\mu\mu}}}{m_{\mu\mu}^2}. \quad (4.5)$$

Pseudo datasets are generated across all event categories from the post-fit background shapes obtained for each type of function in each subcategory, taking into account the uncertainties in the fit parameters as well as their correlations, and injecting a given number of signal events. Signal-plus-background fits are performed on the pseudo datasets using the core-pdf method. The median difference between the measured and injected signal yields, relative to the post-fit uncertainty in the signal yields, gives an estimate of the bias due to the choice of the background model. The bias measured in each BDT subcategory, as well as from pseudo datasets in which the signal is injected simultaneously in all event categories, is smaller than 20% of the post-fit uncertainty on the signal yield. The core-pdf method employed in this analysis yields an improvement in sensitivity of about 10% with respect to the background functions used in the previous result [132], while the number of degrees of freedom in the signal extraction fit is significantly decreased.

4.7.4 Fit procedure

The signal extraction is performed using a maximum-likelihood fit of background-only and signal-plus-background hypotheses to the observed data. The fit is performed simultaneously in all BDT subcategories with the signal strength modifier as a common floating parameter. Figure 4.15 shows the $m_{\mu\mu}$ distributions in each of the ggH subcategories, in which the signal is extracted by performing a binned maximum-likelihood fit using a DCB function to model the signal contribution, while the background is estimated with the core-pdf method.

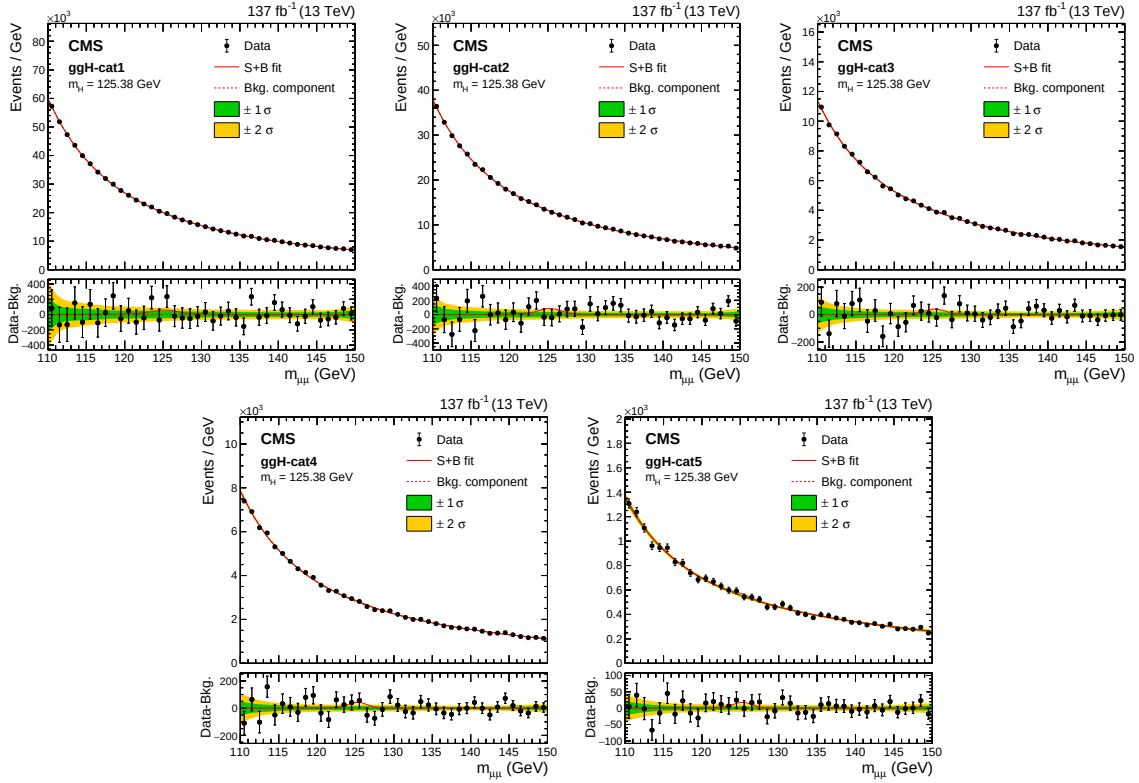


Figure 4.15. Signal-plus-background fits in the ggH subcategories [1].

4.8 $t\bar{t}H$ category

The $t\bar{t}H$ process has the smallest cross section among the targeted Higgs boson production modes at the LHC. However, the presence of a top quark-antiquark pair in addition to the Higgs boson helps to reduce the background to a level that is comparable to the expected signal rate. The top quark decays predominantly into a b quark and a W boson, therefore a sample of events enriched in $t\bar{t}H$ production is selected by requiring the presence of at least two jets passing the loose working point of the DeepCSV b -tagging algorithm [118], or at least one b -tagged jet passing the medium working point.

The $t\bar{t}H$ signal events may contain additional charged leptons, depending on the decay of the top quarks. Events with one or two additional charged leptons in the final state are grouped in the $t\bar{t}H$ leptonic category. An event in the $t\bar{t}H$ leptonic category containing three (four) charged leptons is further required to have the net sum of the lepton electric charges equal to one (zero). In the case of events with more than one pair of oppositely charged muons with $110 < m_{\mu\mu} < 150$ GeV, the pair with the largest dimuon p_T is chosen as the Higgs boson candidate. The invariant mass of each pair of same-flavour, opposite-sign leptons is required to be greater than 12 GeV to suppress backgrounds arising from quarkonium decays. An event is vetoed if it contains a pair of oppositely charged electrons or muons with an invariant mass in the range 81–101 GeV, consistent with the decay of an on-shell Z boson. In contrast, events with exactly two oppositely charged muons with $110 < m_{\mu\mu} < 150$ GeV, no identified electrons, and at least one combination of three jets in the final state with invariant mass (m_{jjj}) between 100 and 300 GeV belong to the $t\bar{t}H$ hadronic category.

The dominant background in the $t\bar{t}H$ hadronic category comes from fully leptonic $t\bar{t}$ decays, while the main backgrounds in the $t\bar{t}H$ leptonic category are the $t\bar{t}Z$ and $t\bar{t}$ processes. In order to obtain an optimal discrimination between the $t\bar{t}H$ signal and the expected backgrounds, BDT-based multivariate discriminants are trained in both the hadronic and leptonic categories.

The input variables are chosen to account for both the kinematic properties of the dimuon system and the properties of the top quark decay products, while ensuring that the BDT outputs remain uncorrelated with $m_{\mu\mu}$. A common set of observables is used as input to the

two BDT discriminants. These include variables that characterize the production and decay of the Higgs boson candidate, namely $p_T^{\mu\mu}$, $y_{\mu\mu}$, ϕ_{CS} , and $\cos\theta_{CS}$. In addition, the η of each of the two muons and the ratio of each muon's p_T to $m_{\mu\mu}$ are also considered. To account for the large hadronic activity in $t\bar{t}H$ signal events, the p_T and η of the three leading jets, the maximum DeepCSV value of jets not overlapping with charged leptons ($\Delta R(\ell, j) > 0.4$), the number of jets, and the scalar (vectorial) p_T sum H_T ($H_T^{\text{miss.}}$) of all identified leptons and jets ($p_T > 25$ GeV, $|\eta| < 2.5$) are included. The $p_T^{\text{miss.}}$ is also considered along with the $\Delta\zeta$ variable, which is defined as the projection of the $\vec{p}_T^{\text{miss.}}$ on the bisector of the dimuon system in the transverse plane. Signal events are weighted during the BDT training with the inverse of the per-event mass resolution, following the same approach used in the ggH categories.

In the $t\bar{t}H$ leptonic category, several additional variables are used in the BDT discriminant that target the kinematic properties of a leptonic top quark decay. These include the azimuthal separation between the Higgs boson candidate and the highest p_T additional charged lepton, the invariant mass formed by the leading additional lepton and the jet with the highest DeepCSV score, and the transverse mass formed by the additional lepton and $E_T^{\text{miss.}}$ in the event. In the $t\bar{t}H$ hadronic category, the resolved hadronic top tagger (RHTT), which combines a kinematic fit and a BDT-based multivariate discriminant, is used to identify top quark decays to three resolved jets. The jet triplet with the highest RHTT score is selected as a hadronic top quark candidate. The corresponding RHTT score is used as input to the BDT discriminant. Furthermore, the p_T of the top quark candidate and the p_T -balance ratio of the top quark and the muon pair are also considered.

Figure 4.16 shows the output of the BDT discriminant in the $t\bar{t}H$ hadronic (left) and leptonic (right) categories.

Based on the BDT output, events in the $t\bar{t}H$ leptonic category are further divided into two subcategories, labelled as “ttHlep-cat1” and “ttHlep-cat2”, corresponding to signal efficiency quantiles of 0–52 and >52%, respectively. Similarly, events in the $t\bar{t}H$ hadronic category are divided into three subcategories labelled “ttHhad-cat1”, “ttHhad-cat2”, and “ttHhad-cat3”, corresponding to signal efficiency quantiles of 0–70, 70–86, and >86%, respectively.

Figure 4.17 shows the $m_{\mu\mu}$ distributions in the $t\bar{t}H$ hadronic and leptonic event categories. The signal is extracted by performing a binned maximum-likelihood fit to these $m_{\mu\mu}$ distributions (bin size of 50 MeV), where signal is modeled using the DCB function (Eq. 4.2) and the background is modeled using a second-order Bernstein polynomial in $t\bar{t}H$ had-cat1 and $t\bar{t}H$ had-cat2, a sum of two exponentials in $t\bar{t}H$ had-cat3, and a single exponential in the $t\bar{t}H$ leptonic event categories. Similarly to how it is done in the ggH category, the potential bias arising from the choice of the background fit function is taken into account by generating pseudo datasets and fitting them with various parametric functions. The corresponding bias is observed to be smaller than 20% of the post-fit uncertainty on the signal yield and is therefore neglected in the signal extraction.

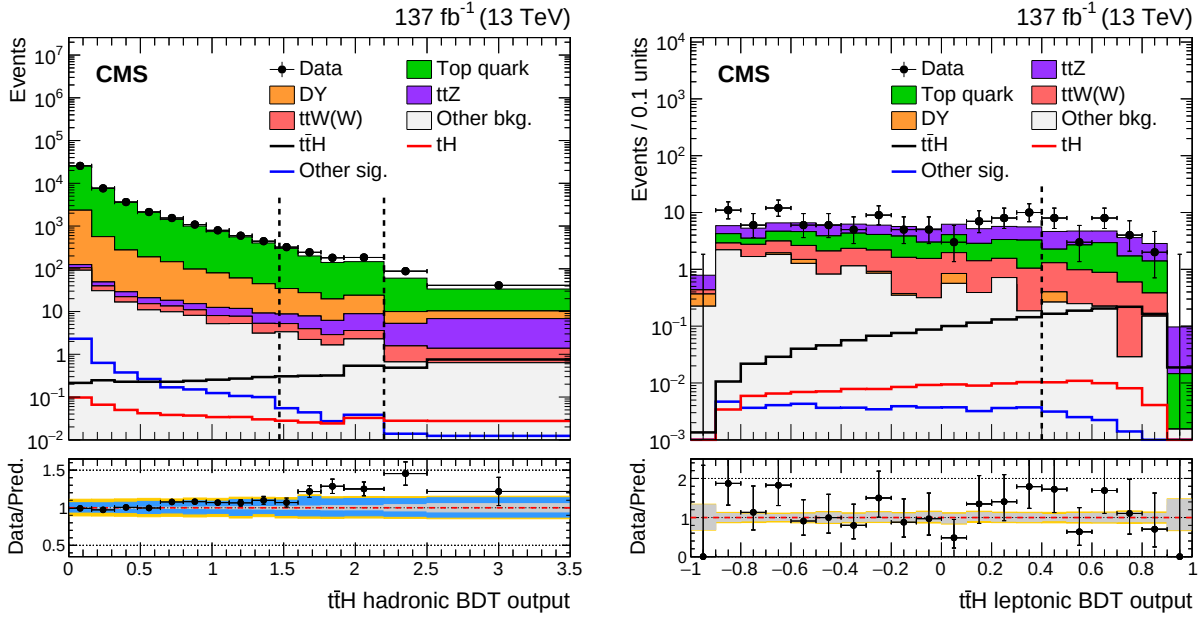


Figure 4.16. BDT score distributions and subcategory boundaries in $t\bar{t}H$ -hadronic (left) and $t\bar{t}H$ -leptonic (right) categories [1].

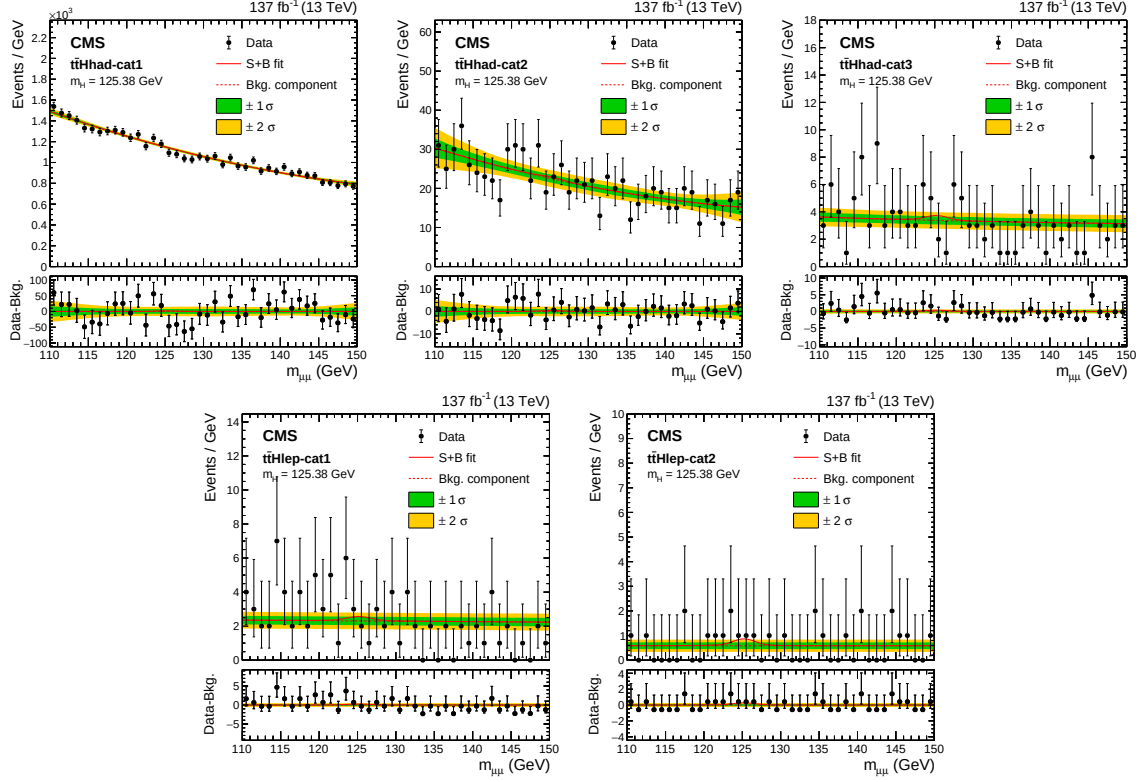


Figure 4.17. Signal-plus-background fits in the $t\bar{t}H$ subcategories [1].

4.9 WH and ZH categories

Events selected for the WH and ZH categories are required to pass the pre-selection criteria and to contain at least one additional charged lepton (electron or muon), which is expected from the leptonic decay of the W or Z boson. The additional muons (electrons) must have $p_T > 20$ GeV, $|\eta| < 2.4$ (2.5), and pass certain isolation and identification requirements with an average efficiency of 95 (90)%. Furthermore, all muons and electrons in these categories are required to pass the medium working point of a multivariate discriminant [136] developed to identify and suppress non-prompt leptons, with a selection efficiency of about 95 (92)% per prompt muon (electron).

Events containing exactly one additional charged lepton belong to the WH category, which targets signal events where the Higgs boson is produced in association with a leptonically decaying W boson. If the additional lepton is a muon, the two pairs of oppositely charged muons are required to have $m_{\mu\mu} > 12$ GeV to suppress background events from

quarkonium decays. Moreover, neither of the two oppositely charged muon pairs can have an invariant mass consistent with m_Z within 10 GeV. Finally, at least one of these two muon pairs must have $m_{\mu\mu}$ in the range 110–150 GeV. If both $m_{\mu\mu}$ pairs satisfy this criterion, the pair with the highest $p_T^{\mu\mu}$ is considered as the Higgs boson candidate. If the additional lepton is an electron, the only requirement imposed is that $110 < m_{\mu\mu} < 150$ GeV.

The ZH category targets signal events where the Higgs boson is produced in association with a Z boson that decays to a pair of electrons or muons. Events in the ZH category are therefore required to contain four charged leptons, with a combined lepton number and electric charge of zero. As in the WH category, the invariant mass of each pair of same-flavour, opposite-sign leptons is required to be greater than 12 GeV. An event is rejected if it does not contain exactly one pair of same-flavour, opposite-sign leptons with invariant mass compatible with the Z boson within 10 (20) GeV for muon (electron) pairs. In addition, each event must contain one oppositely charged muon pair satisfying $110 < m_{\mu\mu} < 150$ GeV. For events with four muons, the muon pair with $m_{\mu\mu}$ closer to m_Z is chosen as the Z boson candidate, while the other muon pair is selected as the Higgs boson candidate.

Two BDT discriminants are trained to discriminate between signal and background events in the WH and ZH categories. The input variables are selected such that the BDT outputs are not significantly correlated with the mass of the Higgs boson candidate. This is required by the chosen analysis strategy, which is analogous to that adopted for the signal extraction in the ggH category. The impact of the $m_{\mu\mu}$ resolution, which evolves as a function of muon p_T and η , is taken into account during the BDT training by applying weights to the simulated signal events that are inversely proportional to the per-event mass resolution.

The BDT discriminant used in the WH category takes as inputs several variables that exploit the kinematic features of the three charged leptons in the event, as well as the p_T^{miss} . These variables include the full kinematic information, apart from the invariant mass, of the dimuon system corresponding to the Higgs boson candidate. In addition, the $\Delta\phi$ and $\Delta\eta$ separations between the additional lepton (ℓ_W) and the Higgs boson candidate, between ℓ_W and both muons from the Higgs boson candidate, and between ℓ_W and $\vec{H}_T^{\text{miss}}$ are considered. The $\vec{H}_T^{\text{miss}}$ is defined as the negative vector p_T sum of all jets in the event with $p_T > 30$ GeV and $|\eta| < 4.7$. Finally, the transverse mass of the combined ℓ_W and $\vec{H}_T^{\text{miss}}$ system, the

flavour of ℓ_W , and the p_T of ℓ_W are added as inputs to the BDT. The particular kinematic properties in signal events of the ℓ_W and $\vec{H}_T^{\text{miss}}$ enable a large suppression of the residual DY background. The BDT discriminant trained in the ZH category considers several input observables constructed from the lepton pair associated with the Z boson decay and the muon pair considered as the Higgs boson candidate. The flavour of the lepton pair associated with the Z boson decay is also included as an input variable.

Figure 4.18 shows the output of the BDT classifiers in the WH (left) and ZH (right) categories. Based on these outputs, events in the WH category are further divided into three subcategories termed “WH-cat1”, “WH-cat2”, and “WH-cat3” corresponding to signal efficiency quantiles of 0–22, 22–70, >70%, respectively. Similarly, events in the ZH category are divided into two subcategories, labelled “ZH-cat1” and “ZH-cat2” corresponding to signal efficiency quantiles of 0–52 and >52%, respectively.

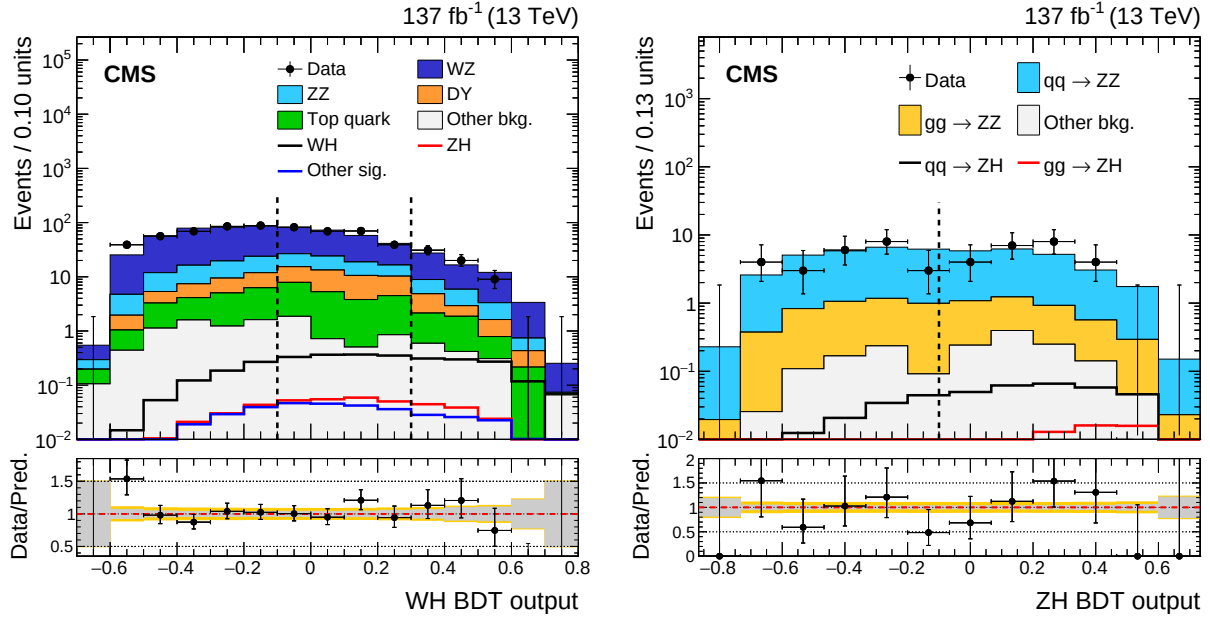


Figure 4.18. BDT score distributions and subcategory boundaries in WH (left) and ZH (right) categories [1].

Figure 4.19 shows the $m_{\mu\mu}$ distributions in the WH and ZH event categories. The signal is extracted via a binned maximum-likelihood fit in each event category, where the signal is modeled with a DCB function (Eq. 4.2) and the background is modeled with the BWZ γ function (Eq. 4.5) in WH-cat1, and the BWZ (Eq. 4.3) function in the remaining subcategories. Similarly to how it is done in the ggH category, the potential bias arising from the choice of the background fit function is taken into account by generating pseudo datasets and fitting them with various parametric functions. The corresponding bias is observed to be smaller than 20% of the post-fit uncertainty on the signal yield and is therefore neglected in the signal extraction.

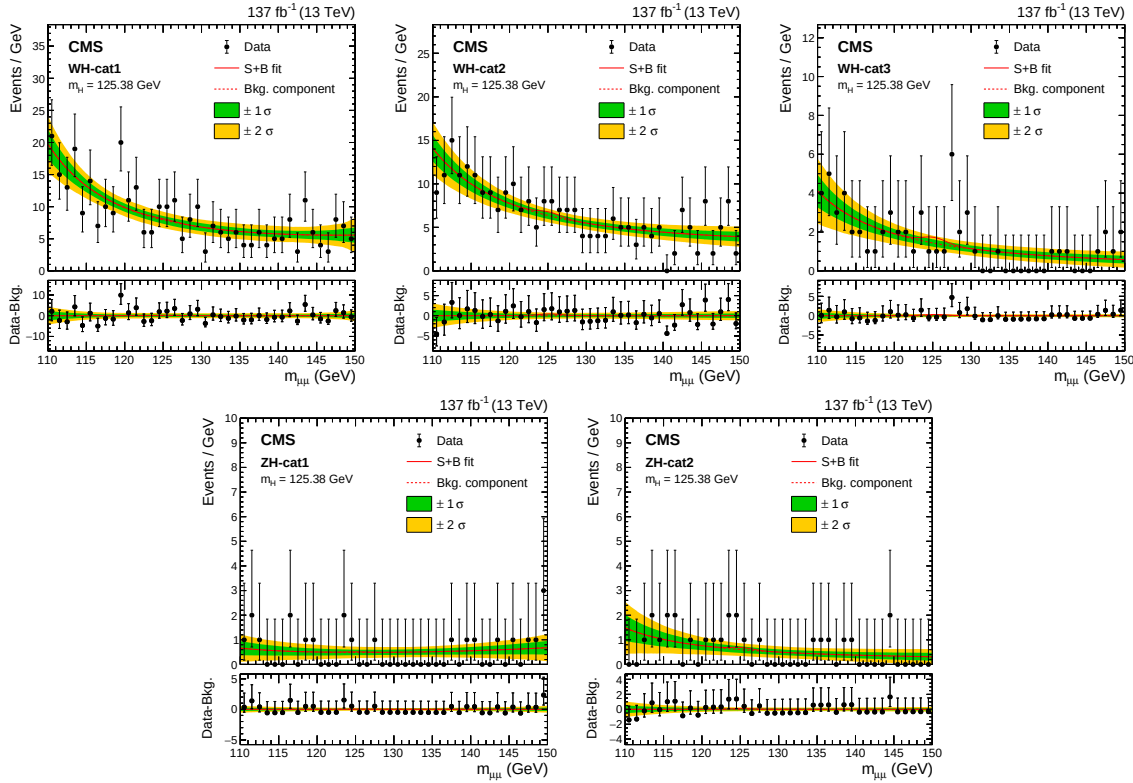


Figure 4.19. Signal-plus-background fits in the VH subcategories [1].

4.10 Combined results

The combined fit in all analysis categories is performed using the signal strength modifier (μ) as a floating parameter. The signal strength modifier is defined as the ratio of the observed $H \rightarrow \mu\mu$ rate to the rate predicted by the SM:

$$\mu = \frac{(\sigma\mathcal{B}(H \rightarrow \mu\mu))_{\text{obs.}}}{(\sigma\mathcal{B}(H \rightarrow \mu\mu))_{\text{SM}}}. \quad (4.6)$$

The relative contributions from the different Higgs boson production modes are fixed to the SM prediction within uncertainties. Confidence intervals on the signal strength are estimated using a profile likelihood ratio test statistic [130], in which systematic uncertainties are modeled as nuisance parameters following a modified frequentist approach [137]. The profile likelihood ratio is defined as

$$q = -2 \ln \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_a)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})},$$

where $\hat{\mu}$ represents the value of the signal strength that maximizes the likelihood $\mathcal{L}$ for the data, while $\hat{\theta}$ and $\hat{\theta}_\mu$ denote the best fit estimate for the nuisance parameters and the estimate for a given fixed value of μ , respectively.

The local p -value quantifies the probability for the background to produce a fluctuation larger than the apparent signal observed in the search region. Figure 4.20 (left) shows the observed local p -value for the combined fit, and for each individual production category, as a function of m_H in a 5 GeV window around the expected Higgs boson mass. The solid markers indicate the mass points for which the observed p -values are computed. Figure 4.20 (right) shows the expected p -values computed for the combined fit, and for each production category, on an Asimov data set [130] generated from the background expectation obtained from the S+B fit with a $m_H = 125.38$ GeV signal injected. The observed p -values as a function m_H are compatible, within the statistical variation, with the expectation for the Higgs boson with $m_H = 125.38$ GeV.

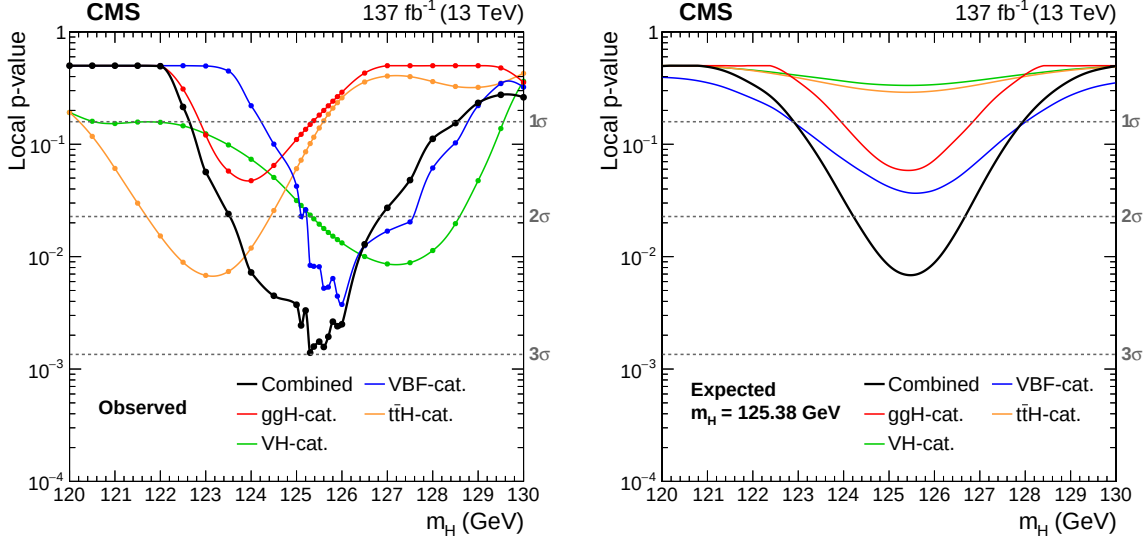


Figure 4.20. Left: observed local p -values as a function of m_H , extracted from the combined fit as well as from each individual production category, are shown. The solid markers indicate the mass points for which the observed p -values are computed. Right: the expected p -values are calculated using the background expectation obtained from the signal-plus-background fit and injecting a signal with $m_H = 125.38$ GeV and $\mu = 1$ [1].

The observed (expected for $\mu = 1$) significance at $m_H = 125.38$ GeV of the incompatibility with the background-only hypothesis is 3.0 (2.5) standard deviations. The 95% CL upper limit (UL) on the signal strength, computed with the asymptotic CLs criterion [130, 138, 139], is also derived from the combined fit performed across all event categories. The observed (expected for $\mu = 0$) UL on μ at 95% CL for $m_H = 125.38$ GeV is 1.9 (0.8).

The best fit signal strength for the Higgs boson with mass of 125.38 GeV, and the corresponding 68% CL interval, is $\hat{\mu} = 1.19^{+0.41}_{-0.40}(\text{stat.})^{+0.17}_{-0.16}(\text{syst.})$. Assuming SM production cross sections for the various modes, the $H \rightarrow \mu\mu$ branching fraction is constrained at 95% CL to be within $0.8 \times 10^{-4} < \mathcal{B}(H \rightarrow \mu\mu) < 4.5 \times 10^{-4}$. The statistical component of the post-fit uncertainty is separated by performing a likelihood scan as a function of μ in which nuisance parameters associated with systematic uncertainties are fixed to their best fit values. The systematic uncertainty component is then taken as the difference in quadrature between the total and the statistical uncertainties. The individual contributions to the uncertainty in the

measured signal strength from experimental uncertainties, the limited size of the simulated samples, and theoretical uncertainties are also evaluated following a similar procedure.

Figure 4.21 reports a summary of the best fit values for the signal strength and the corresponding 68% CL intervals obtained from a profile likelihood scan in each production category. The best fit signal strengths in each production category are consistent with the combined fit result as well as the SM expectation. A likelihood scan is performed in which the four main Higgs boson production mechanisms are associated to either fermion (ggH and $t\bar{t}H$) or vector boson (VBF and VH) couplings. Two signal strength modifiers, denoted as $\mu_{\text{ggH},t\bar{t}H}$ and $\mu_{\text{VBF,VH}}$, are varied independently as unconstrained parameters in the fit. The best fit values for these parameters are $\hat{\mu}_{\text{ggH},t\bar{t}H} = 0.66^{+0.67}_{-0.66}$ and $\hat{\mu}_{\text{VBF,VH}} = 1.84^{+0.89}_{-0.77}$, consistent with the SM expectation.

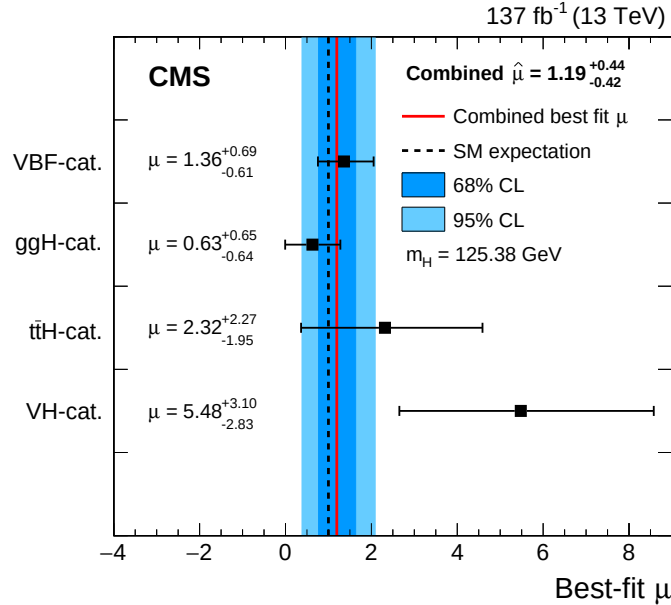


Figure 4.21. Signal strength modifiers measured for $m_H = 125.38$ GeV in each production category (black points) are compared to the result of the combined fit (solid red line) and the SM expectation (dashed grey line) [1].

These results are combined with those obtained from data recorded at centre-of-mass energies of 7 and 8 TeV. The 7+8 TeV search described in Ref. [140] has been updated using the values reported in Ref. [43] for the Higgs boson production cross sections and branching fractions. Systematic uncertainties in the inclusive signal production cross sections and

branching ratios are correlated across the 7, 8, and 13 TeV analyses. Experimental uncertainties affecting the measured properties of the various physics objects (muons, electrons, jets, and b -quark jets), the measurement of the integrated luminosity, and the modeling of the pile-up conditions are assumed to be uncorrelated between the 7+8 and 13 TeV analyses. Table 4.3 reports the observed and expected significances over the background-only expectation at $m_H = 125.38$ GeV and the 95% CL ULs on μ in each production category, as well as for the 13 TeV and the 7+8+13 TeV combined fits. The combination improves, relative to the 13 TeV-only result, both the expected and the observed significance at $m_H = 125.38$ GeV by about 1%. Figure 4.22 shows the observed (solid black) and the expected (dashed black) local p -values derived from the 7+8+13 TeV combined fit as a function of m_H in a 5 GeV window around the expected Higgs boson mass. The expected p -value is computed on an Asimov data set generated from the background expectation obtained from the S+B fit with a $m_H = 125.38$ GeV signal injected. The best fit signal strength, and the corresponding 68% CL interval, obtained from the 7+8+13 TeV combination for the Higgs boson with mass of 125.38 GeV is $1.19^{+0.40}_{-0.39}(\text{stat.})^{+0.15}_{-0.14}(\text{syst.})$.

Table 4.3. Observed and expected significances for the incompatibility with the background-only hypothesis for $m_H = 125.38$ GeV and the corresponding 95% CL upper limits on μ (in the absence of $H \rightarrow \mu\mu$ decays) for each production category, as well as for the 13 TeV and the 7+8+13 TeV combined fits.

Production category	Obs. (exp.) signif.	Obs. (exp.) UL on μ
VBF	2.40 (1.77)	2.57 (1.22)
ggH	0.99 (1.56)	1.77 (1.28)
$t\bar{t}H$	1.20 (0.54)	6.48 (4.20)
VH	2.02 (0.42)	10.8 (5.13)
Combined $\sqrt{s} = 13$ TeV	2.95 (2.46)	1.94 (0.82)
Combined $\sqrt{s} = 7, 8, 13$ TeV	2.98 (2.48)	1.93 (0.81)

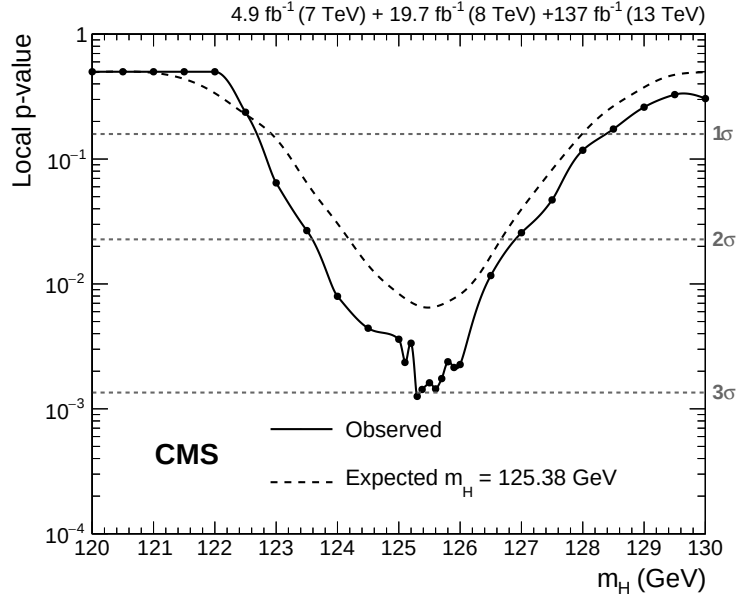


Figure 4.22. Observed (solid black) and expected (dashed black) local p -values as a function of m_H , extracted from the combined fit performed on data recorded at center-of-mass energies of 7, 8, and 13 TeV, are shown. The expected p -values are calculated using the background expectation obtained from the S+B fit and injecting a signal with $m_H = 125.38$ GeV and $\mu = 1$ [1].

The result presented here is the most precise measurement of the $H \rightarrow \mu\mu$ decay rate reported to date, and provides the best constraint of the coupling between the Higgs boson and the muon. The signal strength measured in the $H \rightarrow \mu\mu$ analysis cannot be translated directly into a measurement of the Higgs boson coupling to muons because it is also sensitive to the interactions between the Higgs boson and several SM particles involved in the production processes considered, primarily the top quark and vector boson couplings. These Higgs boson couplings to other particles are constrained by combining the result of this analysis with those presented in Ref. [141]. Under the assumption that there are no new particles contributing to the Higgs boson total width, Higgs boson production and decay rates in each category are expressed in terms of coupling modifiers within the κ -framework [142]. Six free coupling parameters are introduced in the likelihood function (κ_W , κ_Z , κ_t , κ_τ , κ_b , and κ_μ) and are extracted from a simultaneous fit across all event categories.

The best fit value for κ_μ ($\kappa_\mu = 1.07$), as well as those for the other couplings, are found to be compatible with the SM prediction. The 68% CL interval for the κ_μ parameter

is $0.85 < \kappa_\mu < 1.29$, corresponding to the total symmetrized uncertainty of approximately 22%, while the 95% CL interval is $0.59 < \kappa_\mu < 1.50$.

In the SM, the Yukawa coupling between the Higgs boson and the fermions (λ_F) is proportional to the fermion mass (m_F), while the coupling to weak bosons (g_V) is proportional to the square of the vector boson masses (m_V). The results from the κ -framework fit can therefore be translated in terms of reduced coupling strength modifiers, defined as $y_V = \sqrt{\kappa_V} m_V/\nu$ for weak bosons and $y_F = \kappa_F m_F/\nu$ for fermions, where ν is the vacuum expectation value of the Higgs field of 246.22 GeV [120]. Figure 4.23 shows the best fit estimates for the six reduced coupling strength modifiers as a function of particle mass, where lepton, vector boson, and quark masses are taken from Ref. [120]. The compatibility between the measured coupling strength modifiers and their SM expectation is derived from the $-2 \Delta \ln(\mathcal{L})$ separation between the best fit and an alternative one, performed by fixing the six coupling modifiers to the SM prediction ($\kappa_W = \kappa_Z = \kappa_t = \kappa_\tau = \kappa_b = \kappa_\mu = 1$), yielding a p -value of 44%.

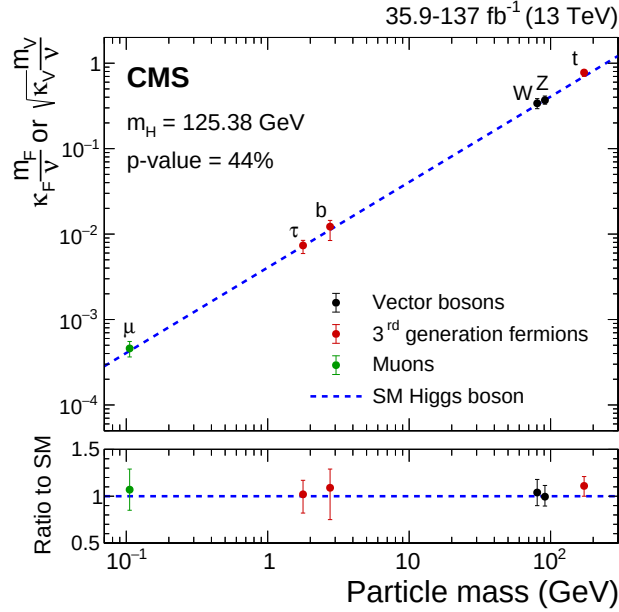


Figure 4.23. The best fit estimates for the reduced coupling modifiers extracted for fermions and weak bosons from the resolved κ -framework compared to their corresponding prediction from the SM. The error bars represent 68% CL intervals for the measured parameters. In the lower panel, the ratios of the measured coupling modifiers values to their SM predictions are shown [1].

5. NEW DEVELOPMENTS FOR THE $H \rightarrow \mu\mu$ ANALYSIS STRATEGY

In this chapter, the key aspects of the MC template-based analysis strategy used in the VBF category of the Run 2 $H \rightarrow \mu\mu$ search are reviewed, and possible directions of improvement are suggested. Section 5.1 describes the studies related to the object selection and categorization in the VBF phase space, while Section 5.2 focuses on the MVA methods used in the MC-based analysis approach. The suggested ideas for the analysis optimization can be utilized in the next iteration of the $H \rightarrow \mu\mu$ analysis with the data collected during Run 3 of the LHC.

5.1 Studies for the MC template-based strategy

The analysis category targeting the VBF production mode is based on the topology of the events of interest. Such events are required to have a pair of muons and a pair of jets satisfying the requirements mentioned in Section 4.6. The sensitivity to the signal events is extracted from the kinematics of the muon pair and kinematics of the jet pair, as well as their relative position and topology in an event.

The dimuon kinematics provides a separating power between signal and background because the $H \rightarrow \mu\mu$ signal forms a peak in the dimuon mass spectrum, while the background processes follow a smoothly falling distribution. The jet pair kinematics characterizes the distinctive topology of the VBF $H \rightarrow \mu\mu$ events, allowing to discard a significant portion of the background.

The studies exploring the possible directions of improving the sensitivity by using the kinematic features of the dimuon and dijet pairs are presented in this section. The studies described in Sections 5.1.2 and 5.1.3 use deep neural networks (DNN) as an optimization tool to extract useful information; more details about DNN design and parameters can be found in Appendix B.

5.1.1 Dimuon mass resolution

The shape of the invariant mass distribution of the $H \rightarrow \mu\mu$ signal is completely defined by the dimuon mass resolution of the detector (several GeV), as the width of the $H \rightarrow \mu\mu$ decay is much smaller (several MeV). In the data-driven analysis strategy, used in the ggH , $t\bar{t}H$ and VH categories, the MVA models are required to be uncorrelated with the dimuon mass to avoid the mass sculpting, because the mass distribution is used as an observable. In the VBF category, decorrelation of the DNN with dimuon mass is not required. On the contrary, the dimuon invariant mass is used as one of the inputs to the DNN, providing the highest discriminating power among the event properties. Thus, the detector resolution has a direct impact on the analysis sensitivity, as the separating power of the dimuon mass distribution would be higher for a narrower signal.

The invariant mass resolution can be computed on a per-event basis, using the uncertainties of the transverse momenta of the muons:

$$\sigma_m/m_{\mu\mu} \approx \frac{1}{2} \sqrt{\left(\frac{\sigma_{p_T}}{p_T}\right)_{\mu_1}^2 + \left(\frac{\sigma_{p_T}}{p_T}\right)_{\mu_2}^2}, \quad (5.1)$$

where we neglect the uncertainties on the angular coordinates of muons, as they are typically insignificant in comparison to the p_T uncertainties.

As shown in Figure 5.1, the events with good mass resolution are signal-like for the events with $m_{\mu\mu} \approx 125$ GeV and background-like in the mass sidebands, while for the events with a worse mass resolution this correlation is inverted. Although the discriminating power of the mass resolution is low, it does provide a small improvement to the DNN performance when used as an input in addition to $m_{\mu\mu}$, as illustrated in Figure 5.8.

The distribution of the per-event mass resolution derived using the uncertainties of the muon momenta may be different between the simulation and the observed data due to imperfections in muon reconstruction. This may lead to a bias in the DNN sensitivity if the mass resolution is used as a DNN input feature. Therefore, one must ensure that the per-event resolution is properly calibrated before using it in training. As shown in Figure 5.2,

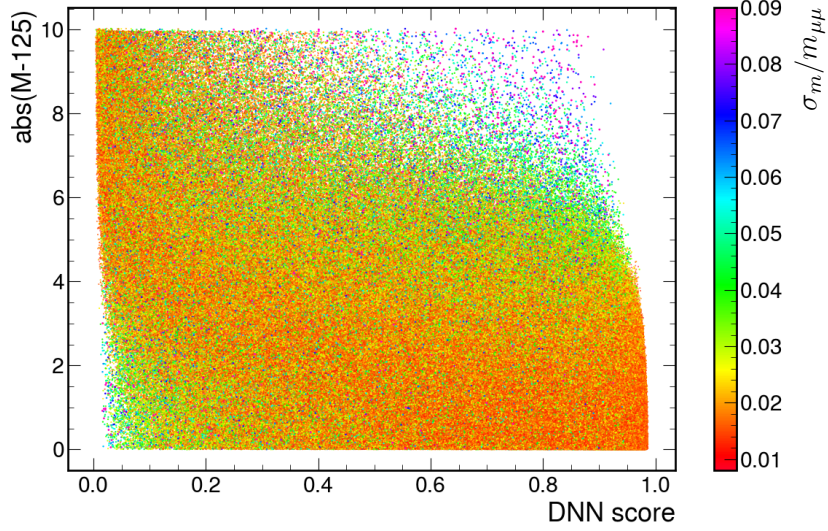


Figure 5.1. Correlation between the dimuon mass and the VBF DNN score for VBF $H \rightarrow \mu\mu$ events. The color represents the relative event-by-event mass resolution $\sigma_m/m_{\mu\mu}$.

in the Run 2 analysis, a difference of about 10% was observed between data and MC in the per-event mass resolution distributions, while the overall level of agreement was adequate.

The effect of a potential improvement of the analysis sensitivity due to an improved dimuon mass resolution was evaluated by altering the invariant mass values for the signal events during evaluation of the DNN score templates. The nominal dimuon mass resolution as a function of the muon pseudorapidity¹ is shown in Figure 5.3 (right). For an $x\%$ improvement, the dimuon invariant mass was updated as follows:

$$m_{\mu\mu} \rightarrow 125 \text{ GeV} + (m_{\mu\mu} - 125 \text{ GeV}) * (100\% - x\%),$$

such that for $x = 0$ the value of the mass would remain unchanged, and for $x = 100$ all signal events would have dimuon mass 125 GeV.

The values of the event-by-event mass resolution are not updated, as their correlation with the DNN score is non-trivial, and naïve scaling does not lead to a realistic estimate. The

¹↑ The largest absolute pseudorapidity between the two muons is used, since the invariant mass resolution is dominated by the muon with the worse p_T resolution. The p_T resolution deteriorates at higher $|\eta|$ values, due to a decrease in the magnetic field strength.

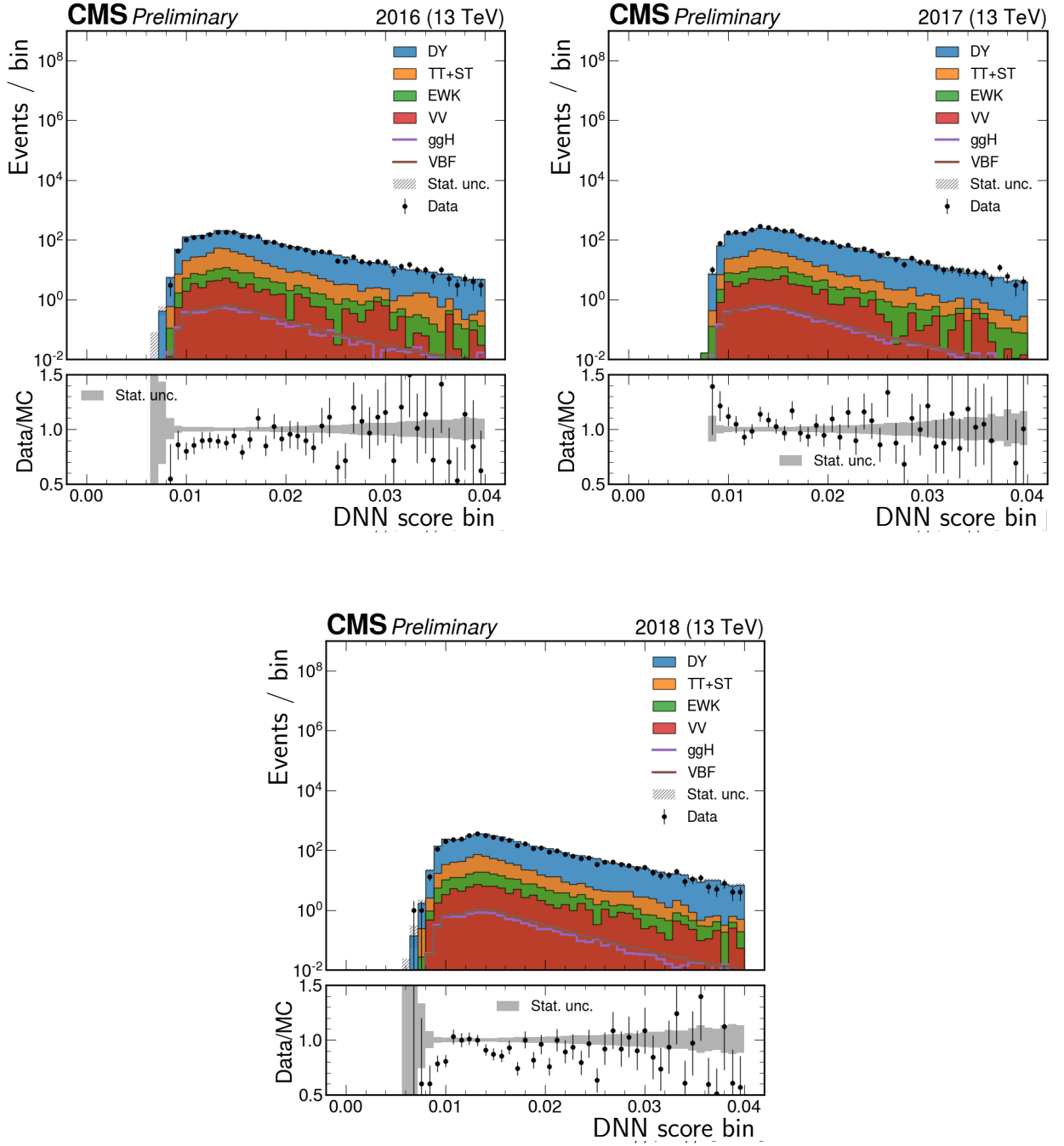


Figure 5.2. Relative per-event dimuon mass resolution distribution in 2016 (top left), 2017 (top right), 2018 (bottom).

updated mass values were used to evaluate the DNN model described in Section 5.2.1 for the signal datasets, without retraining the network. Then, the updated signal templates were used to estimate the expected significance. The background templates remained unchanged in all cases.

Figure 5.3 (left) shows the evolution of the expected significance for the mass resolution improvements from 0 to 100%. The signal significance in the VBF category scales linearly with the mass resolution improvement, until the improvement reaches approximately 70%. The change of behavior for larger values can be attributed to artifacts due to DNN score binning and the fact that the DNN was not re-trained with a narrow signal spectrum. In the linear part of the plot, the slope indicates that 1% improvement in mass resolution would lead to $\approx 0.5\%$ improvement in signal significance.

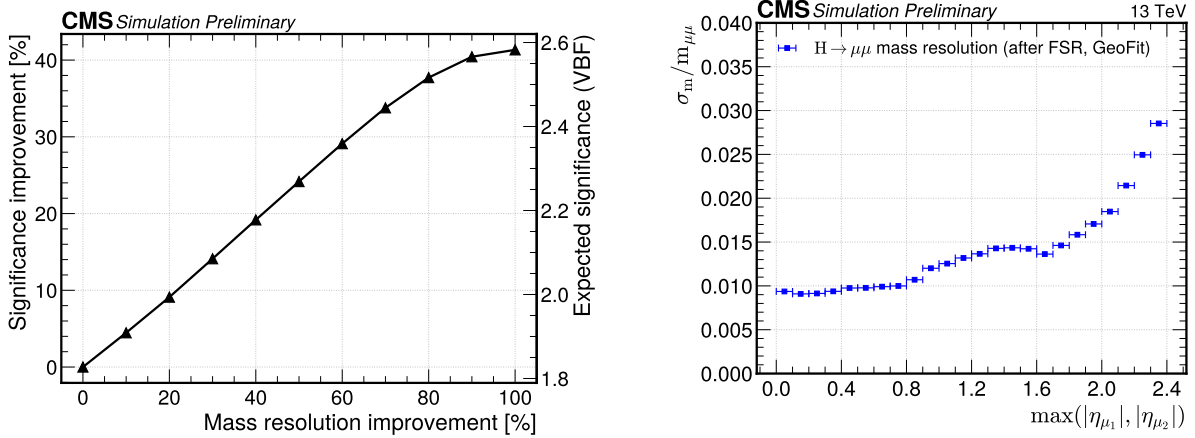


Figure 5.3. Left: expected significance in the VBF channel as a function of the mass resolution improvement. The improvement is estimated by varying only the $m_{\mu\mu}$, σ_m , and $\sigma_m/m_{\mu\mu}$ inputs of the VBF DNN. Resolution improvement of 100% corresponds to $m_{\mu\mu} = 125$ GeV for all ggH and VBF signal events. Right: dimuon mass resolution for $H \rightarrow \mu\mu$ events as a function of the largest absolute pseudorapidity of the two muons.

5.1.2 Jet pair selection

As described in Section 4.6, the events in the category targeting the VBF $H \rightarrow \mu\mu$ decay channel in the Run 2 analysis were required to contain a pair of jets passing certain kinematic selections ($m_{jj} > 400$ GeV, $\Delta\eta_{jj} > 2.5$, $p_T^{j_1} > 35$ GeV). After an initial pre-selection of jets, the pair of interest was selected as the two leading- p_T jets. While in most of the events the method of selecting the jet pair does not make a difference because only one jet pair being present, there is a considerable share of events in which there is more than one way to choose a jet pair.

Jet pair selection methods

In most of the considered MC datasets, there is no more than two jets generated at the matrix element level. Any additional generated jets are produced at the parton shower simulation step, which is known to model the kinematics of the additional jets imperfectly. Therefore, the agreement between data and MC is expected to be worse for those jets and the discriminating power of their kinematics is reduced. The reconstructed jets in an event originate from different sources; some of them may correspond to the generated jets, while others stem from the pile-up, or mis-reconstructed from the detector noise.

The performance of the jet pair selection criteria can be assessed using geometrical matching between the reconstructed and generated jets. Although this approach does not distinguish between the jets produced at different generator steps, it can provide an insight into the effectiveness of the pile-up and detector noise rejection for different selection methods.

Tables 5.1 and 5.2 show the distribution of the number of reconstructed jets in the VBF $H \rightarrow \mu\mu$ dataset, as well as the distribution by number of reconstructed jets that are matched with a generated jet within $\Delta R < 0.4$. In approximately 17% of events, there are at least 3 reconstructed jets present, therefore there is more than one way to select a jet pair. However, in about a half of those events, all of the reconstructed jets are matched to a generated jet, therefore a “correct” pair cannot be easily chosen.

Table 5.1. The distribution of the events in the total Run 2 VBF $H \rightarrow \mu\mu$ dataset by the number of selected reconstructed jets, as well as by the number of those jets matched with a generated jet.

		Number of RECO jets matched to GEN						Total
		0	1	2	3	4	≥ 5	
# of RECO jets	0	90353	-	-	-	-	-	90353
	1	127620	409986	-	-	-	-	537606
	2	10906	353992	535786	-	-	-	900684
	3	1167	29349	112164	105783	-	-	248463
	4	134	2910	10047	22339	17842	-	53272
	≥ 5	43	396	1090	2328	5127	3674	12658
Total		230223	796633	659087	130450	22969	3674	1843036

Table 5.2. The distribution of the events in the in the total Run 2 VBF $H \rightarrow \mu\mu$ dataset by the number of selected reconstructed jets, as well as by the number of those jets matched with a generated jet.

		Number of RECO jets matched to GEN						Total
		0	1	2	3	4	≥ 5	
# of RECO jets	0	4.9%	-	-	-	-	-	4.9%
	1	6.9%	22.2%	-	-	-	-	29.2%
	2	0.6%	19.2%	29.1%	-	-	-	48.9%
	3	0.1%	1.6%	6.1%	5.7%	-	-	13.5%
	4	0.0%	0.2%	0.5%	1.2%	1.0%	-	2.9%
	≥ 5	0.0%	0.0%	0.1%	0.1%	0.3%	0.2%	0.7%
Total		12.5%	43.2%	35.8%	7.1%	1.2%	0.2%	100%

In order to understand how much can be gained from improving the jet selection, we focus on the events that have

- at least 3 reconstructed jets;
- at least 2 of them matched to generated jets;
- but not all of the reconstructed jets are matched to the generated jets.

For such events, there is more than one way to select a pair, and there is at least one “true” pair of generated jets which can be used to estimate the accuracy of the choice. According to Table 5.2, these events account for approximately 8.3% of all VBF events. Multiple methods to select a jet pair are considered:

- Select two leading- p_T jets (used in the Run 2 analysis)
- Select 2nd and 3rd leading- p_T jets (skip the jet with highest p_T)
- Select two jets with highest invariant mass m_{jj}
- Select two jets with highest angular separation $\Delta\eta_{jj}$

The accuracy of these methods is shown in Table 5.3. The jet pair is considered correctly selected if both of the selected jets are matched to a generated jet. Three of the considered methods show only $\approx 30\%$ accuracy, while simply skipping the leading- p_T jet increases the accuracy up to 63%. This may be due to the fact that in many of the events with more than two jets, the leading- p_T jet originates from pile-up.

Table 5.3. Accuracy of jet pair selection methods evaluated using events with at least 3 reconstructed jets, of which at least two are matched to generated jets, but not all of them.

Pair selection method	Accuracy
1 st and 2 nd leading- p_T jets	30.6%
2 nd and 3 rd leading- p_T jets	63.3%
Highest m_{jj}	25.1%
Highest $\Delta\eta_{jj}$	32.0%

DNN-based jet pair selection

As there is room for improvement of the jet selection accuracy, a new method was developed based on machine learning. The events with at least three reconstructed jets, of which at least two, but not all were matched to the generated jets, were used to train a DNN. The inputs of the network (60 input features) included the kinematics of up to five selected jets, as well as the dimuon pair. The full set of input features is listed below:

- dimuon system: p_T , η , ϕ ;
- for each of the two muons: p_T , η , ϕ ;

- for up to 5 jets: p_T , η , ϕ , quark-gluon likelihood score (QGL) [128], the working point of the pile-up rejection criteria (PUID), and the working point of the jet selection criteria (JetID);
- $\Delta\eta$ and $\Delta\phi$ for each muon-jet pair;
- number of jets in the event.

The network featured a fully connected multi-layer architecture with three hidden layers containing 64, 32, and 16 nodes, respectively, and an output layer with 10 nodes corresponding to all possible pairs of the five leading jets.

The truth value for a given output node was set to 1 if both of the jets in the pair corresponding to that node are matched with generated jets, and to 0 otherwise. Since more than one pair of generated jets may exist in the event, multiple output nodes may be “true” at the same time.

The input dataset contained the full Run 2 VBF $H \rightarrow \mu\mu$ sample, and was split into three folds, with 50% of events used for training, 25% for validation, and 25% for evaluation. Four networks were trained, each one with a different dataset splitting, in order to make use of all events (see Section B.1.5).

The cross-entropy loss function was minimized using the Adam optimizer [110]. The training was performed with batch size 4096 and lasted up to 10 epochs; the model with the lowest loss was saved and used for final evaluation.

At the evaluation stage, the output node with the highest value was selected as the “DNN prediction”. The pair selection accuracy achieved with this setup was 81.0%, which significantly exceeds the accuracy of other methods shown in Table 5.3. However, as the jet pair selection method affects only a fraction of events, the overall impact on the analysis is not expected to be large.

The input features with the highest impact on the DNN output were found to be the transverse momenta (p_T) of all considered objects (jets, muons, and the dimuon pair), the PUID working points of the considered jets, as well as the angular separation between the jets and the leading muon. The DNN prediction for all considered events was either to pick the two highest- p_T jets, or the 2nd and 3rd jets by p_T .

Results

The DNN-based jet pair selection was implemented in the analysis and propagated to the templates in the VBF category. The categorization cuts on m_{jj} and $\Delta\eta_{jj}$ were preserved, which resulted in overall decrease in yields in the VBF category, as compared to Run 2, due to the dijet mass being on average lower than before. The DNN in the VBF category was retrained using the same settings as in other studies (see Section 5.2.1), and the binning by the DNN score was defined to preserve the same VBF yields in the high-score bins as in the Run 2.

The combined expected significance for the updated jet pair selection was calculated to be 1.67σ , which is lower than the Run 2 significance of 1.8σ . The reason for that could be a non-negligible impact of the events where one or both of the jets come from pile-up. In these events, the dimuon invariant mass, which is one of the VBF DNN inputs, is still a powerful feature for signal-background separation. Therefore, such events contribute to the sensitivity even if the distributions of the jet pair kinematics are not as sensitive.

The studies described in this section indicate that in most of the events with ≥ 2 generated jets, these jets can be found among the three highest- p_T reconstructed jets. At the same time, using all three of the leading jets in the DNN training will necessarily result in a degraded data vs. MC agreement, since at least one of these jets is reconstructed from either pile-up, detector noise, or parton shower. A possible strategy can be to take into consideration multiple jets before the event categorization, use a pre-trained method to select a pair of jets that is believed to provide the best signal-background discrimination power and the best agreement between data and MC, and use that pair for the DNN training.

5.1.3 Event categorization in 2-jet phase space

In the Run 2 analysis, the boundary between the categories targeting ggH and VBF production modes was defined such that the events with a jet pair with invariant mass $m_{jj} > 400$ GeV, angular separation $\Delta\eta_{jj} > 2.5$ and leading jet $p_T > 35$ GeV would be included into the VBF category. The choice of this boundary was motivated by the fact that the VBF $H \rightarrow \mu\mu$ events are characterized by a jet pair with high invariant mass and angular

separation. To estimate the efficiency of the categorization, we have calculated the share of ggH and VBF $H \rightarrow \mu\mu$ events that end up in the category targeting the respective production mode. With the Run 2 approach, the categorization efficiency for ggH was 86.7%, while for VBF it was 62.8% (the values are computed only for events with at least two reconstructed jets).

In an attempt to construct a category containing the VBF events with a cleaner signature, a new categorization method was developed. The approach implements a DNN classifier trained to distinguish between ggH and VBF events containing two jets based on the kinematics of those jets. The jet pair selection is the same as in the original Run 2 analysis, without using the developments considered in the previous section.

To validate the approach, the DNN was first tested with only three inputs — m_{jj} , $\Delta\eta_{jj}$, p_T , and a trivial architecture with one hidden layer containing three nodes. Then, the same setup was trained with a deep 3-layer architecture containing 64, 32, and 16 nodes, respectively. Finally, the 3-layer architecture was tested with an extended set of input features, which included 15 variables:

- p_T , η , ϕ , QGL for each of the two jets;
- p_T , η , ϕ , mass, $\Delta\eta$, $\Delta\phi$ of the dijet system;
- number of jets in the event.

The ROC curves for these setups are shown in Figure 5.4. The setups with 3 input features provided similar efficiencies to the Run 2 cuts, while the training with the full set of features has improved the efficiency. With the new approach, for a fixed Run 2 VBF efficiency value (63%), the ggH efficiency increased by 5% from 87% to 91%. For a fixed ggH efficiency (87%), the VBF efficiency increased by 10% from 63% to 69%.

To estimate the effect of the new categorization on the sensitivity in the VBF category, the category boundary was defined as a cut on the MVA categorizer score, such that the VBF efficiency is the same as in the Run 2 analysis. This way, the VBF $H \rightarrow \mu\mu$ yield in the VBF category is preserved, which allows to use the same definition of binning for the templates (in

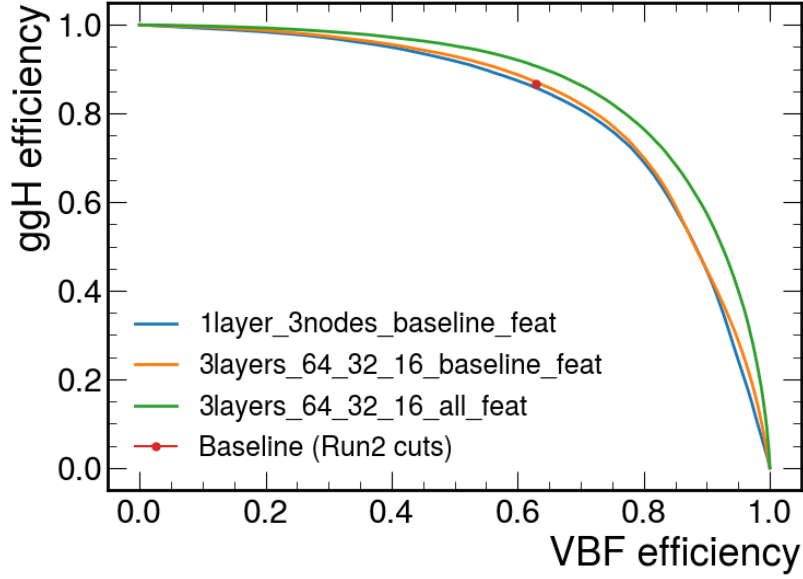


Figure 5.4. ggH vs. VBF categorization efficiency in Run 2, compared to the new MVA-based method. Red: Run 2 efficiencies. Blue and orange: DNN setup with 3 input features. Green: DNN setup with 15 input features.

Run 2, the bins were defined based on VBF yields). The VBF DNN was retrained using the same settings as in other studies (see Section 5.2.1), and the new templates were produced.

The expected significance with this setup was calculated to be 1.57σ , which is considerably lower than the significance in Run 2 (1.8σ). The primary reasons for decreased sensitivity are:

- Decreased ggH yield in the high-score bins; the contribution of the ggH signal was non-negligible, even though the category was meant to target the VBF production mode.
- Increased background yields, due to MVA categorizer being trained without taking any background information into consideration.

Lack of improvement in the VBF category sensitivity for a better ggH vs. VBF separation indicates that the approach to categorization should be revised: the category in which the template-based analysis strategy is used should target not only the VBF production mode,

but more generally, the sensitive part of the 2-jet phase space, which includes both VBF $H \rightarrow \mu\mu$, as well as ggH $H \rightarrow \mu\mu$ signal with two additional jets.

At the same time, the trade-off between improving sensitivity in the template-based category and potential loss of sensitivity in the remaining part of the 2-jet phase space should be considered. Optimization of the combined sensitivity is not a trivial task, due to the fact that the data-driven strategy requires the category boundaries to be uncorrelated with dimuon mass, while the sensitivity in the template-based strategy is highly correlated with the dimuon mass.

5.1.4 Breakdown of signal sensitivity in VBF phase space

Although, in principle, both of the studies described in Sections 5.1.2 and 5.1.3 were aimed to achieve a purer VBF signature, they did not lead to an improvement in the sensitivity of the VBF analysis category. One of the reasons for this stems from the fact that the signal significance has non-negligible contributions not only from the “clean” VBF $H \rightarrow \mu\mu$ events, but also from the ggH $H \rightarrow \mu\mu$ events, and from the VBF events with at least one pile-up jet misidentified as a jet from the VBF process. The DNN score distributions in Figure 5.5 correspond to the setup reproducing the VBF category sensitivity in Run 2, but the VBF $H \rightarrow \mu\mu$ signal is split based on the number of the jets that were matched to generated jets.

The expected significance for the individual components of the signal, as well as for their combination is shown in Table 5.4. The contribution from the ggH signal accounts for approximately 20% of the total signal sensitivity, while the contribution of the VBF events with one or more misidentified jets accounts for approximately 30%.

Table 5.4. Contribution of different components of the $H \rightarrow \mu\mu$ signal in the VBF phase space.

Process			Total
VBF 0–1 matched GEN jets	0.56σ	1.45σ	1.82σ
VBF 2 matched GEN jets	0.93σ		
ggH	0.37σ		

5.1.5 Template-based strategy in 1-jet phase space

The possibility of extracting the signal using a template-based method was tested for events containing exactly one reconstructed jet passing the general jet selection. The jet in the event may originate from different sources:

- signal or background process itself, like in VBF $H \rightarrow \mu\mu$ or in EWK $Z \rightarrow \ell\ell$, if the second jet is out of the detector acceptance range, or if the reconstruction of the jet failed
- initial state radiation of a gluon (ISR)
- pile-up interactions
- detector noise

The kinematic differences between jets of different sources may provide a certain sensitivity for signal-background separation. The list of input features was taken from the “VBF” category, and cleaned from the variables that require the presence of two jets:

- Dimuon variables: $m_{\mu\mu}$, $\log m_{\mu\mu}$, $p_T^{\mu\mu}$, $\eta_{\mu\mu}$, σ_m , $\sigma_m/m_{\mu\mu}$, $\cos\theta_{CS}$, ϕ_{CS}
- Jet variables: p_T^j , η_j , ϕ_j , QGL_j , $\Delta\eta_{\mu\mu,j}$
- Other variables: $N_{\text{jets}}^{\text{soft}}(p_T > 5 \text{ GeV})$, $H_T^{\text{soft}}(p_T > 2 \text{ GeV})$, year (2016/2017/2018).

The DNN architecture and training settings were similar to the DNN used for the studies in the VBF category. The binning of the output score was defined such that each of 70 highest bins contained 1 ggH event (after applying all analysis weights), and the rest of the events were grouped in one low-score bin. The score distribution for different data-taking periods is shown in Figure 5.6.

The expected signal significance computed with these templates is 1.142σ . While in the VBF category the S/B ratio in the highest-purity bin was approaching 1, in the 1-jet category, the best S/B ratio was of the order of 1/100. To fully evaluate the performance of the template-based method in the 1-jet category, it should be compared with the data-driven method in the same phase space.

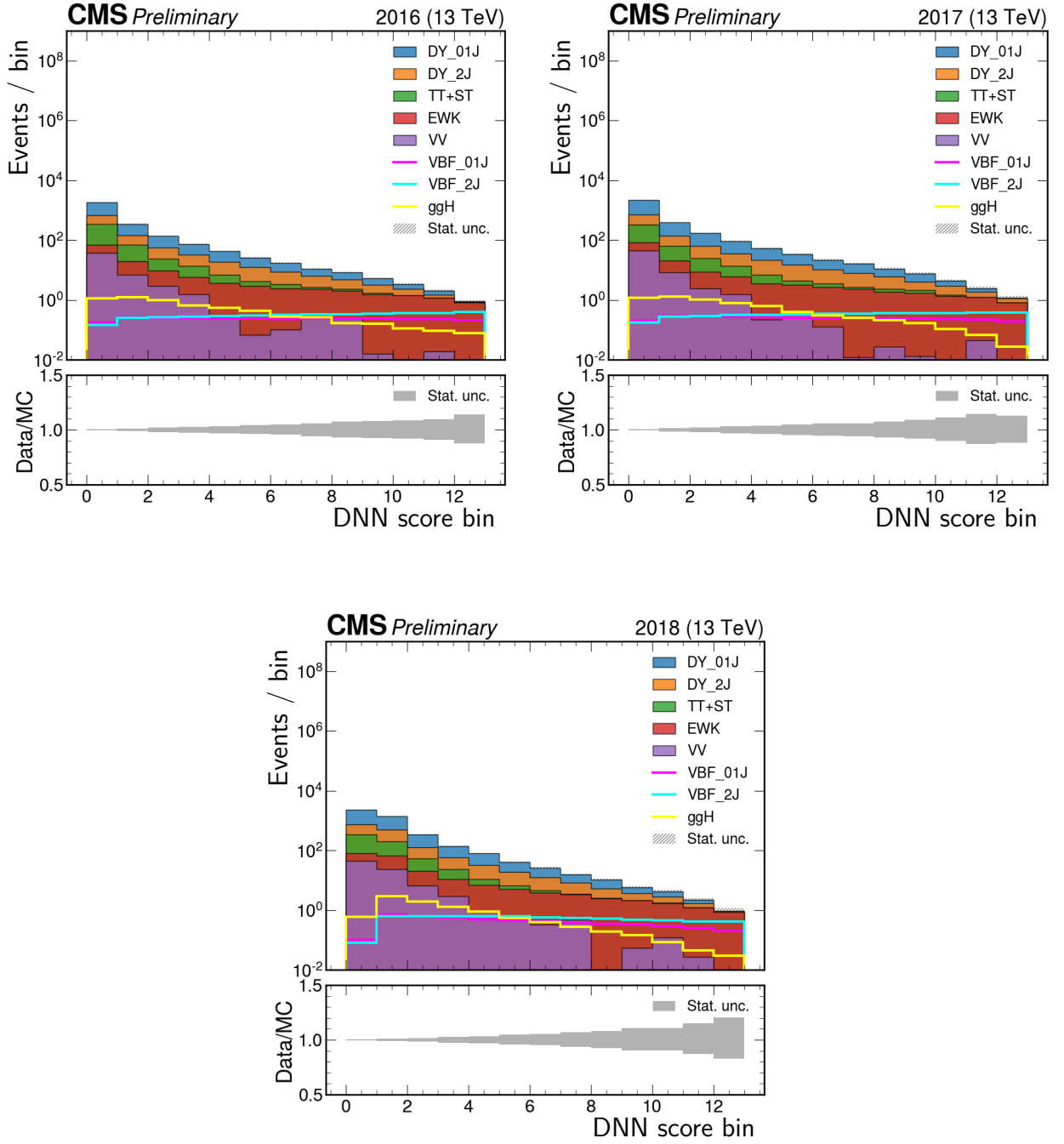


Figure 5.5. DNN score distribution in the VBF channel. Top left: 2016, top right: 2017, bottom: 2018.

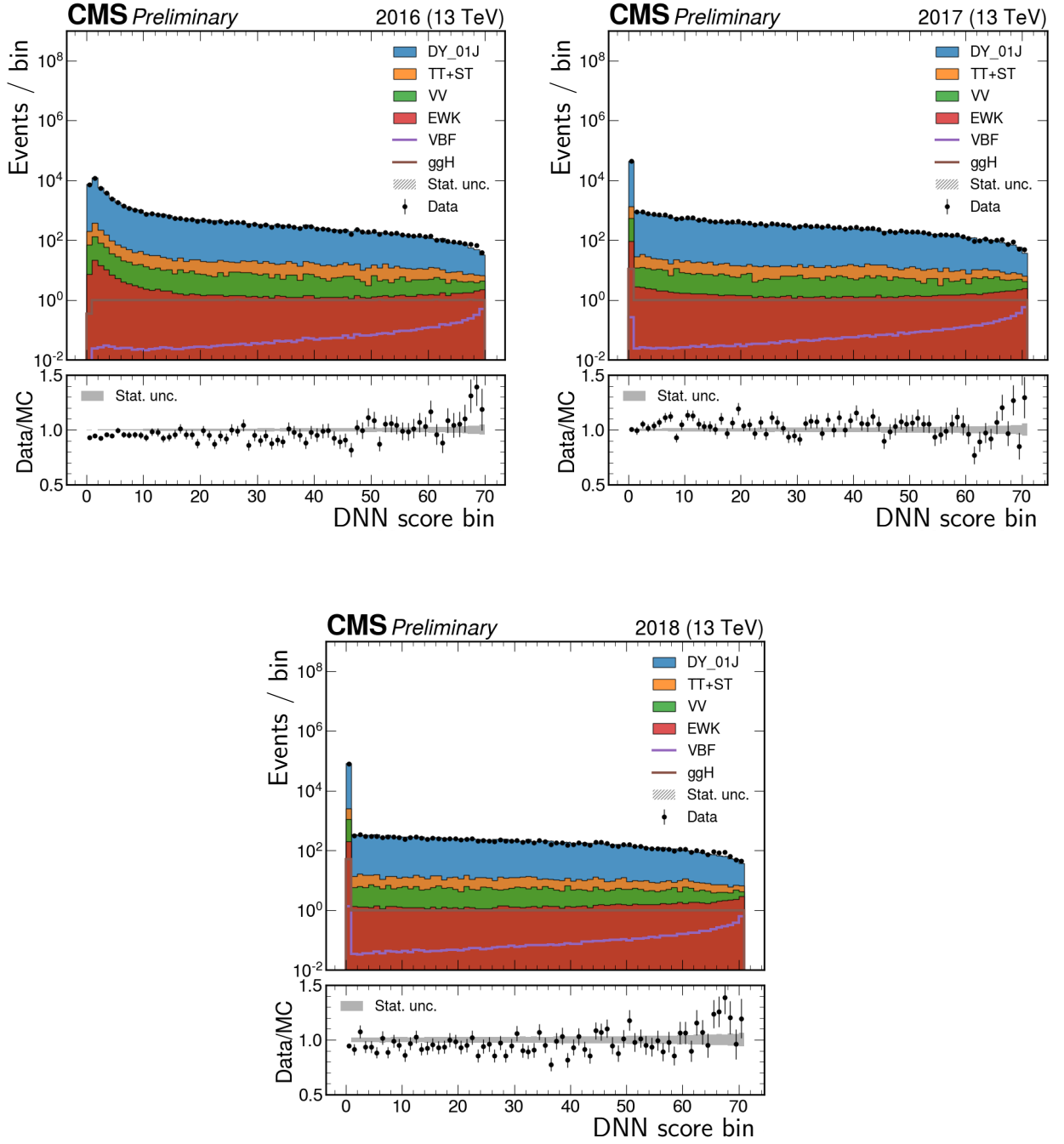


Figure 5.6. DNN score distribution in 1-jet category for 2016, 2017 and 2018 data-taking periods.

5.2 MVA methods in the VBF phase space

The central part of the MC-based analysis strategy is a multivariate method used to combine the sensitivity from multiple event properties into a one-dimensional observable. In Run 2, a DNN was used for this purpose, while a boosted decision tree (BDT) was considered as an alternative.

This section describes a DNN setup that was used to reproduce the Run 2 sensitivity, as well as studies aiming to identify the key aspects that define the performance of the MVA models.

5.2.1 Baseline DNN setup

DNN architecture and training settings

The expected pre-fit significance of the VBF channel in the Run 2 analysis was 1.77σ , taking into account all considered systematic uncertainties. In order to speed up the studies described in this chapter, several sources of uncertainties were omitted, such as jet energy scale (JES) and resolution (JER) variations, and PDF uncertainties. Without these uncertainties, the expected pre-fit significance is increased to 1.8σ , and this value is used as a baseline for the studies.

This section describes the DNN setup that allows to achieve 1.8σ sensitivity. Instead of re-implementing the Run 2 architecture which featured five subnetworks trained with different datasets and input features, our setup contains a single fully connected network with three hidden layers.

The DNN is implemented using PYTORCH package [143]. The input to the network contains 25 features, which are normalized to be in the same range by subtracting the mean value and dividing by the standard deviation:

- Dimuon system kinematics: $m_{\mu\mu}$, p_T , $\log p_T$, η
- Angles in the Collins-Soper frame [129]: $\cos\theta_{CS}$, ϕ_{CS}
- Event-by-event dimuon mass resolution: σ_m and $\sigma_m/m_{\mu\mu}$

- For each of the two selected jets: p_T , η , ϕ , QGL
- Dijet system kinematics: m_{jj} , $\log m_{jj}$, $\Delta\eta_{jj}$
- p_T -balance ratio: $R_{p_T} = \frac{|\vec{p}_T^{jj} + \vec{p}_T^{\mu\mu}|}{|\vec{p}_T^{j1}| + |\vec{p}_T^{j2}| + |\vec{p}_T^{\mu\mu}|}$
- Zeppenfeld variable [127]: $z^* = \frac{y^*}{|y_{j1} - y_{j2}|}$, where $y^* = y_{\mu\mu} - \frac{y_{j1} + y_{j2}}{2}$
- Angular separation between dimuon pair and the closest jet: $\min\Delta\eta_{\mu\mu,j}$
- Soft activity variables: $N_{\text{jets}}^{\text{soft}}(p_T > 5\text{GeV})$, $H_T^{\text{soft}}(p_T > 2\text{GeV})$
- Year (2016/2017/2018)

The three hidden layers contain 128, 64, and 32 nodes, respectively, with batch normalization and tanh activation applied after each layer. Randomly selected 20% of the weights are dropped at each training iteration to prevent overfitting. The output layer consists of a single node with a sigmoid function [109] applied to its value. The truth value for a given event is 1 for signal and 0 for background events. The binary cross-entropy (BCE) loss function is used to update the network weights during back-propagation, and the minimization of the loss is done using the Adam optimizer [110] with the learning rate parameter 10^{-3} .

Two main backgrounds (Drell–Yan, EWK $Z \rightarrow \ell\ell$) and two signals (VBF and ggH)² are used for training; the VBF signal is a combination of three MC datasets (see details in 5.2.2). The dataset is split into training (50%), validation (25%), and evaluation (25%) folds using the k -fold approach described in Section B.1.5.

The batch size for the training is set to 65536, and the training is performed for 100 epochs for each of the four folds (the number of iterations in one epoch is equal to dataset size divided by the batch size). The best model is saved as the one providing the highest expected significance evaluated on the validation fold (this approach is discussed in more detail in Section B.1.6).

The output of the DNN score is converted to templates (histograms) using the binning that exactly replicates the per-bin yields of the VBF $H \rightarrow \mu\mu$ signal in the Run 2 analysis.

²↑It is worth noting that the ggH signal was not used for the DNN training used in the Run 2 analysis [1].

This is done to exclude any effect of the binning procedure on the synchronization of sensitivity between the two implementations. The DNN score templates for each year are shown in Figure 5.7; the yields for all processes for these templates are synchronized with the Run 2 templates to under 1% precision.

The combined expected pre-fit significance obtained with the described setup is 1.82σ , which closely reproduces the pre-fit significance in the Run 2 analysis for the same phase space (1.78σ). The agreement of the post-fit sensitivity with the Run 2 result has also been confirmed, the expected post-fit significance is 1.80σ in both cases. In the studies described in this chapter only the pre-fit significance is used as a figure of merit, assuming that most of the general conclusions would hold for the post-fit results as well.

Importance of input features

The impact of individual input features on the separating power of the DNN output score has been studied to identify the most sensitive variables. Three methods, illustrated in Figure 5.8 were used to estimate the feature importance:

1. Use the distribution of a single input feature as an observable to compute the significance achievable without MVA. The reference significance value for this method is a simple counting experiment where all events are grouped into a single bin. In the VBF category for the full Run 2 dataset (137 fb^{-1}) the significance of the simple counting experiment is equal to $\approx 0.52\sigma$. The significance achieved if the most sensitive feature, the dimuon mass, is used as an observable is $\approx 0.87\sigma$.
2. Estimate the loss of sensitivity if the contribution of a given feature is suppressed in the DNN output. This is implemented by shuffling the values for a given feature between signal and background datasets, such that the distributions of values of this feature would remain the same and not bias the DNN output, but no separating power could be extracted from it. The reference value for this method was taken as the significance achieved with nominal DNN setup with all features included (the significance was lower than nominal 1.8σ because only one model was evaluated on all events, instead of a proper 4-fold procedure described in Section 5.2.1). The shuffling of the dimuon mass

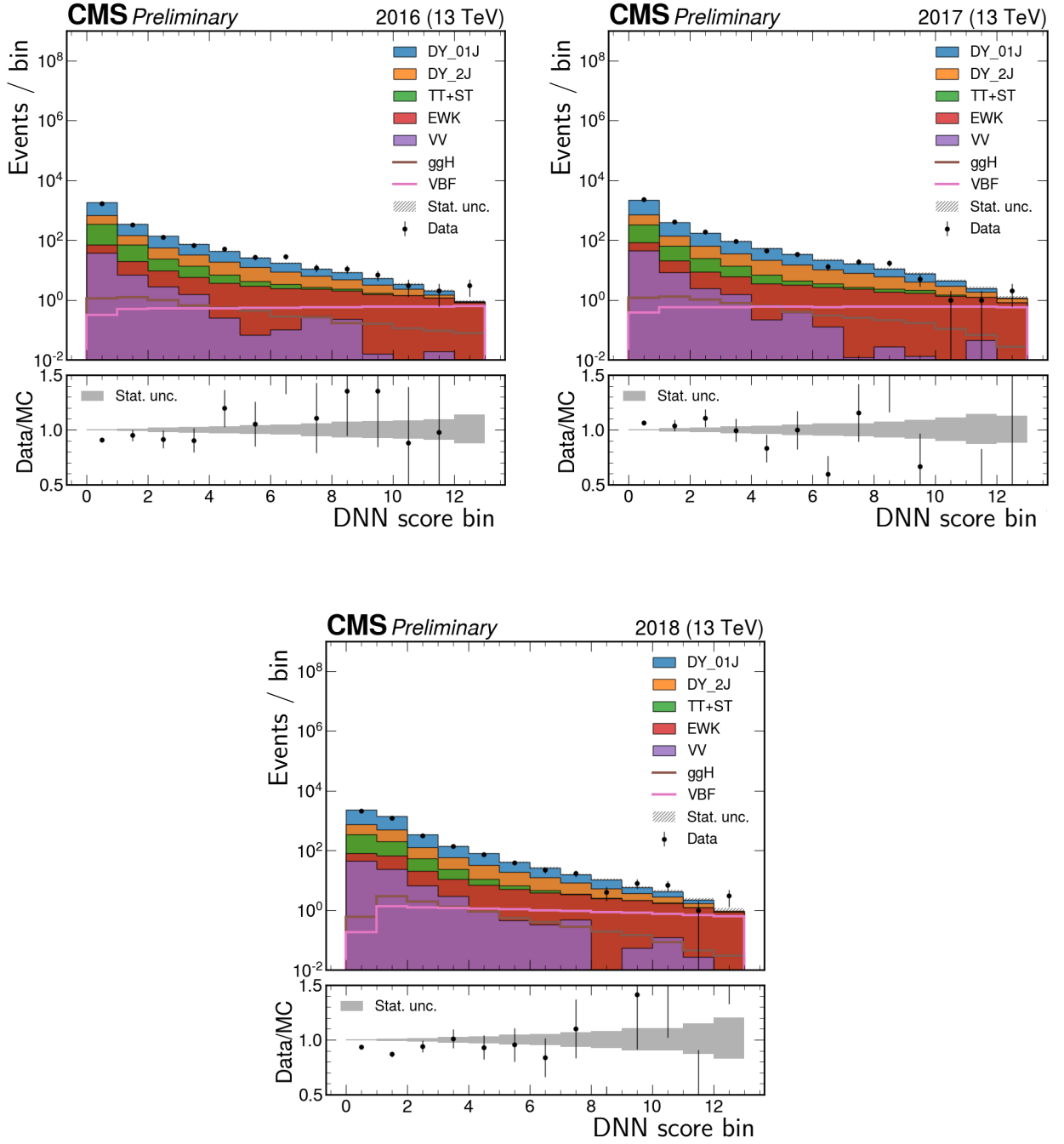


Figure 5.7. DNN score distribution in the VBF channel for the baseline setup replicating the Run 2 sensitivity.

showed the largest impact, lowering the expected significance of the DNN score by $\sim 50\%$.

3. Estimate the feature importance by evaluating the mean value of the gradients at the input layer of the DNN for the VBF $H \rightarrow \mu\mu$ dataset.

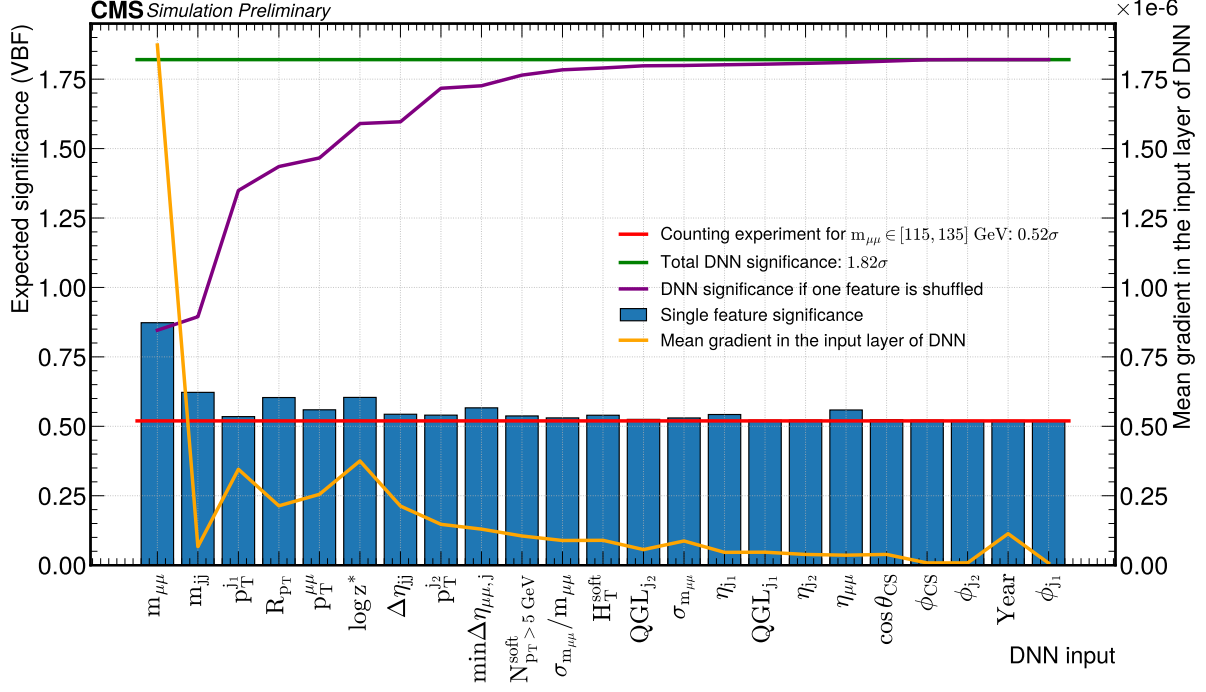


Figure 5.8. Relative importance of DNN input features estimated as their individual sensitivity, a loss of overall sensitivity if impact of a given feature is removed by shuffling its values between signal and background events, and an average gradient at the DNN input node corresponding to the given feature.

All three methods showed consistent results, highlighting that the main contributions to the analysis sensitivity are provided by:

- dimuon mass and dijet mass;
- momenta of dimuon system, jets, and correlation between them (R_{pT});
- angular information of the dijet system ($\Delta\eta$, Zeppenfeld variable).

The low-level angular information (η , ϕ of dimuon pair and jets) does not provide a significant sensitivity neither explicitly (when used as an observable outside of DNN), nor implicitly (when used as DNN inputs together with other low-level variables).

5.2.2 Training datasets

For training of the DNN in the VBF category, the simulation of four processes was used, including two main signal processes, and two backgrounds. These datasets were taken from 2016, 2017 and 2018 MC campaigns, and combined together for training. To increase the signal statistics, three VBF datasets were used, with differences in generators. However, only the `powheg-pythia_dipole` dataset was used to evaluate significance during training, and later to create the DNN score templates, while other datasets were used to estimate the uncertainty related to the parton shower modeling. The VBF $H \rightarrow \mu\mu$ datasets were produced in the next-to-leading order precision in QCD, which means that in addition to the two quark jets in the VBF process, at most one gluon jet was generated at the matrix element level, while other generated jets, if any, were added at the parton shower generation stage.

Each of the signal datasets contained around 1 million events per data-taking year at the generation step, out of which around 240 thousand events were selected for training, making the total size of the VBF $H \rightarrow \mu\mu$ training dataset $\approx 240 \cdot 10^3 \cdot (3 \text{ years}) \cdot (3 \text{ datasets}) \approx 2$ million events.

- Signal:

- VBF $H \rightarrow \mu\mu$

- VBFHToMuMu_M-125_TuneCP5_PSweights_13TeV_powheg_pythia8

- VBFHToMuMu_M-125_TuneCP5_13TeV-powheg-herwigpp

- VBFHToMuMu_M-125_TuneCP5_13TeV-powheg-pythia_dipole

- ggH $H \rightarrow \mu\mu$

- GluGluHToMuMu_M125_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8

- Background:

- Drell–Yan
DYJetsToLL_M-105To160_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8
DYJetsToLL_M-105To160_VBFFilter_TuneCUETP8M1_13TeV-amcatnloFXFX
-pythia8
- EWK $Z \rightarrow \ell\ell$
EWK_LLJJ_MLL_105-160_SM_5f_LO_TuneEEC5_13TeV-madgraph-herwigpp
_corrected

The differences in the generators used for different signal datasets were found to have an effect on DNN performance due to differences in the kinematics and multiplicity of the parton shower jets in the datasets with different generator settings. To illustrate this, the training was repeated with identical settings, but using only one VBF dataset each time. In all cases, the `powheg_pythia_dipole` dataset was used to evaluate significance and produce templates. The left plot in Figure 5.9 (left) shows the evolution of significance computed on the validation fold over 10 thousand training iterations.

As expected, the validation significance is higher when the generator settings are the same for both training and validation datasets. However, since it is unknown which of the signal datasets provides the most accurate description of the jet multiplicity and kinematics in the $H \rightarrow \mu\mu$ decays, the differences in the DNN performance should not be used to conclude that one of the datasets is preferable over the others.

Instead, the training with a combination of different signal datasets is considered (as shown with the blue line in Figure 5.9 (left)), showing a consistent improvement in performance as compared to the trainings with individual datasets. This indicates that regardless of the generator settings used in the training dataset(s), the performance can still be improved by increasing the size of the training dataset.

The dependence of the training process on the amount of signal data was also checked. Figure 5.9 (right) shows the training with varied size of the VBF $H \rightarrow \mu\mu$ dataset, from 50% to 100% of the number of VBF events in the baseline training setup. It can be seen that the performance of the DNN improves on average with more training data, however, there is no evidence that the maximum achieved significance is higher for a larger signal dataset.

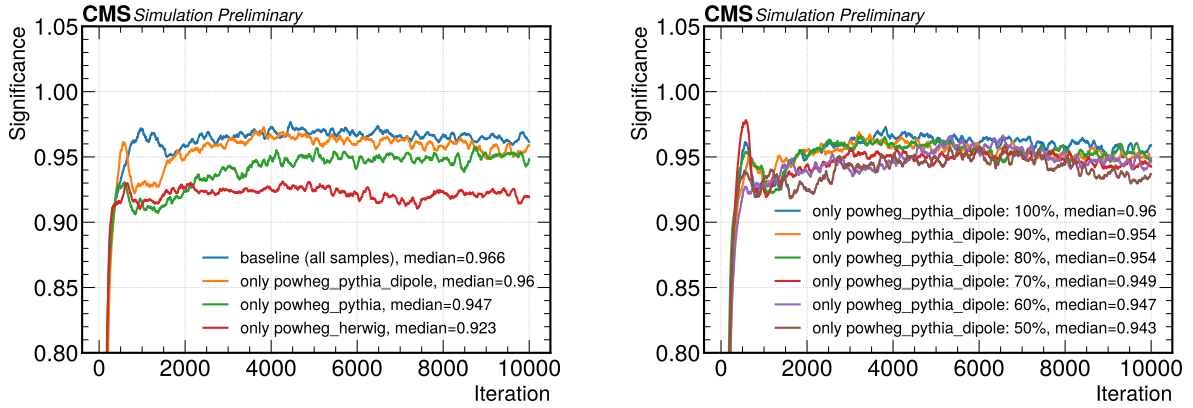


Figure 5.9. Left: significance evaluated on the validation fold for different VBF $H \rightarrow \mu\mu$ datasets and their combination. Right: significance for trainings with variable amount of events in `powheg_pythia_dipole` dataset (other VBF datasets are not used).

In order to not compromise the DNN performance in Run 3, the number of signal events used in training should be at least as large as the number of signal events in Run 2. Thus, at least 3 million signal events should be generated per data-taking year.

The differences between the signal datasets are taken into account in the VBF $H \rightarrow \mu\mu$ analysis by assigning a systematic uncertainty associated with the differences between the signal templates produced using the datasets with different generator settings. This uncertainty effectively accounts for the differences in the DNN performance shown in Figure 5.9 (left), and also suppresses possible improvements of fine-tuning the DNN hyperparameters. In order to decouple the uncertainty on the signal model from the DNN model setup, in the Run 3 $H \rightarrow \mu\mu$ analysis the MVA models should be designed to be less sensitive to systematic variations.

5.2.3 Loss functions

As mentioned in Section B.1.1, a DNN score with statistical power similar to that of the likelihood ratio, in principle, can be obtained by using standard loss functions like cross-entropy and MSE. However, as shown in Section B.1.6, using the cross-entropy loss function does not result in a monotonically increasing significance when the loss is monotonically

decreased, therefore a longer training does not lead to an improvement. It would be beneficial if a loss function was optimizing the significance directly, so that a longer training would lead to a higher level of the signal-background discrimination. We have tested the performance of different loss functions in the DNN used for the $H \rightarrow \mu\mu$ analysis, along with two custom loss functions designed with particle physics analyses in mind (the definitions of these loss functions are given in Section B.1.4):

- Binary cross-entropy (BCE)
- Mean squared error (MSE)
- “Significance loss” [144]
- “Likelihood loss” [145]

The plot in Figure 5.10 shows the evolution of the expected significance for these loss functions for a large number of training iterations. The performance of BCE and MSE losses is very similar, while the setups with “significance loss” and “likelihood loss” do not converge to a high value of the validation significance.

Hence, at the moment there is no evidence that a significant improvement to the analysis sensitivity can be expected from using a different loss function instead of binary cross-entropy.

5.3 Summary

In order to achieve the highest possible sensitivity in the $H \rightarrow \mu\mu$ analysis using Run 3 data, it is important to make use of the lessons learned from the analysis strategy and performance in Run 2.

The gain in sensitivity in the Run 2 analysis, as compared to previous searches for $H \rightarrow \mu\mu$, was achieved by isolating a high-purity region of the VBF $H \rightarrow \mu\mu$ phase space, and using a simulation-based background estimation method instead of a data-driven one. This method allowed to fully exploit the separating power of the event kinematics, primarily the dimuon invariant mass, since the DNN model was not required to be uncorrelated with $m_{\mu\mu}$. Fitting the DNN score has the disadvantage that the bump in the invariant mass distribution is not

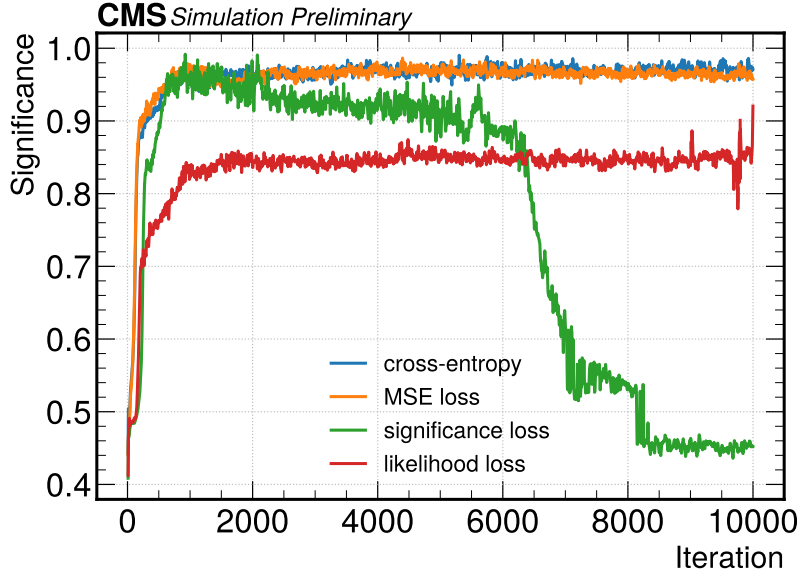


Figure 5.10. Evolution of the expected significance evaluated for the validation fold for different loss functions.

visible. An alternative approach to estimate the background is to perform a parametric fit of the sidebands using MC simulation rather than data, and use the dimuon mass spectrum as an observable instead of the DNN score. However, this approach was not considered in the scope of this thesis.

There are multiple directions in which the analysis strategy can be improved in the Run 3 analysis.

A direct improvement to the sensitivity of the MC template-based strategy can be achieved by enhancing the discriminating power of the DNN input features. For instance, an improvement in the dimuon invariant mass resolution would lead to a greater separation power of the dimuon mass, while an improved jet selection would provide an additional sensitivity to the jet-related training features.

A more flexible approach to the jet pair selection may help to reduce the effect of the resolution loss due to pile-up. This can be achieved by considering more than just the two leading jets in the event selection, and using a multivariate approach to select the jet pair that provides the highest signal-background discrimination power. However, since the parton

shower modeling is imperfect in the MC simulation, the agreement between data and MC needs to be carefully controlled when considering more than 2 jets.

The category boundaries between the regions of the 2-jet phase space in which different analysis strategies are used, can be optimized in such a way that the combined sensitivity of both approaches is maximized. However, we suggest that the category in which the MC template-based method is used should be interpreted not specifically as a category targeting the VBF production mode, but more generally, as a high-purity region of the 2-jet phase space. In this category, a non-negligible contribution is expected from “clean” VBF $H \rightarrow \mu\mu$ events, as well as from the ggH $H \rightarrow \mu\mu$ events, and from VBF events in which one or both jets originate from pile-up, mis-reconstruction, or parton shower.

It has been found that the discriminating power of an observable obtained using a DNN classifier is not very sensitive to the DNN architecture: the performance of a relatively simple three-layer model was found to be similar to that of a complex model used in the Run 2 analysis. The more important aspects of the DNN training were found to be the method of the model selection during training, as well as the size and composition of the signal training dataset. Nevertheless, the optimization of the DNN architecture and other hyperparameters would certainly be beneficial at the fine-tuning stage. The choice of the parton shower model was the main source of systematic uncertainty in the VBF channel of the Run 2 analysis. The DNN performance was found to be dependent on the choice of the signal VBF dataset, which indicates that at least part of the DNN sensitivity (2–3%) is learned from the kinematics of the parton shower jets. Therefore, in the current design, fine-tuning the DNN training settings is not expected to lead to an improvement of the analysis sensitivity, as most of the sensitivity gain would be covered by the parton shower modeling uncertainty. In order to decouple the DNN performance from this uncertainty in the Run 3 analysis, the DNN model should be designed to be insensitive to the parton shower kinematics.

The developments in the directions suggested in this chapter will continue along with the data-taking during the Run 3, and with a sufficient increase in sensitivity, it will be possible to achieve a discovery-level significance of the $H \rightarrow \mu\mu$ signal with the combination of Run 1, Run 2 and Run 3 data.

6. PROSPECTS FOR THE PRECISE MEASUREMENT OF THE HIGGS BOSON PROPERTIES IN THE $H \rightarrow \mu\mu$ DECAY CHANNEL AT THE HL-LHC

According to the planned LHC timeline (see Figure 2.5), after several years of data-taking at $\sqrt{s} = 13.6$ TeV, the LHC will be upgraded to the High-Luminosity LHC (HL-LHC). The HL-LHC will extend the LHC program to the first half of the 2040's, conducting proton-proton collisions at the center-of-mass energy of $\sqrt{s} = 14$ TeV. With the instantaneous luminosity increased by a factor of 10 from the LHC design value of $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, the HL-LHC is expected to produce a large dataset corresponding to an integrated luminosity of 3000 fb^{-1} for each of the ATLAS and CMS experiments. At the same time, the CMS detector will undergo the Phase-2 upgrade [146], affecting most of its subsystems and significantly improving the data-taking performance.

In this chapter, the expected precision for measurements of the Higgs boson properties in the $H \rightarrow \mu\mu$ channel at the HL-LHC with an integrated luminosity of up to 3000 fb^{-1} and a center-of-mass energy of 14 TeV is presented. The precision is estimated by extrapolating the Run 2 results obtained at a center-of-mass energy of 13 TeV and corresponding to an integrated luminosity of 137 fb^{-1} [1]. The extrapolation takes into account the increased center-of-mass energy, as well as the changes to the CMS detector geometry and performance after the Phase-2 upgrade.

6.1 Introduction

The first projections for the $H \rightarrow \mu\mu$ search sensitivity for $L = 3000 \text{ fb}^{-1}$ [147] have been made ahead of the Higgs boson discovery, when the mass of the Higgs boson was not yet known. A generator-level analysis targeting the ggH and VBF Higgs production modes predicted the significance at 3000 fb^{-1} to be 7.1 (7.9) standard deviations, with a total uncertainty on the signal strength of 13(14)% for $m_H = 120(130)$ GeV.

Upon discovery of the Higgs boson in 2012, the projections were revisited for the Snowmass 2013 project [148, 149]. The extrapolation was based on the Run 1 $H \rightarrow \mu\mu$ analy-

sis [150, 151], which used a strategy containing a 0/1-jet category targeting the ggH production mode, and a 2-jet category targeting the VBF production mode. The ggH category was split into 12 subcategories by muon p_T and η , and the VBF category was split into 3 subcategories based on jet pair kinematics. In all categories, a strategy based on parametric fits was used to extract the signal strength modifier. The results of the Run 1 analysis were extrapolated by scaling the event yields to 3000 fb^{-1} , as well as by taking into account the change in cross sections from 8 to 14 TeV. Two uncertainty scaling scenarios have been considered: in scenario 1 (S1), the systematic uncertainties have been kept to their Run 1 values, and the statistical uncertainties have been scaled with integrated luminosity as $1/\sqrt{L}$; in scenario 2 (S2), both systematic and statistical uncertainties have been scaled with integrated luminosity. The projected uncertainty on the signal strength was estimated to be 20 (14)%, and the uncertainty on the coupling modifier κ_μ was estimated to be 8 (7.5)% for S1 (S2), respectively. An accompanying study [152] also estimated the projected significance for 3000 fb^{-1} to be 7.9σ .

The projections for the $H \rightarrow \mu\mu$ sensitivity have been updated in the LHC Yellow Report in 2018 [153] after analyzing the data collected at $\sqrt{s} = 13 \text{ TeV}$ during the 2016 data-taking period, corresponding to an integrated luminosity of 35.9 fb^{-1} . The updated analysis strategy [132] benefited from a new categorization based on a multivariate score constructed using a boosted decision tree (BDT), as well as additional categorization by pseudorapidity. In addition to scaling the event yields according to integrated luminosity and changes in cross sections from 13 to 14 TeV, the improvement in mass resolution due to the Phase-2 upgrade of the CMS tracker [154] has been considered. Again, two uncertainty rescaling scenarios have been considered: fixing the systematic uncertainties to their 2016 values (S1), and rescaling them as $1/\sqrt{L}$ until the “floor” values are reached (S2). A detailed description of the uncertainty scenarios are provided in Section 6.3.4 and in Ref. [155]. The projected uncertainty on the signal strength was estimated to be 12.8 (9.6)%, and the uncertainty on the coupling modifier κ_μ was estimated to be 6.7 (5.0)% for S1 (S2), respectively. The uncertainties are dominated by the statistical component. The 2013 and 2018 projections for the $H \rightarrow \mu\mu$ sensitivity are summarized in Table 6.1.

Table 6.1. Previous projections made for $L = 3000 \text{ fb}^{-1}$.

Projection for 14 TeV, 3000 fb^{-1}		Snowmass 2013	LHC Yellow Report 2018
Expected significance		7.9σ	–
Uncertainty on signal strength	S1	20%	12.8%
	S2	14%	9.6%
Uncertainty on κ_μ	S1	8%	6.7%
	S2	7.5%	5.0%

6.2 The CMS Phase-2 detector

After the end of Run 3, the CMS detector will be substantially upgraded in order to fully exploit the physics potential offered by the increase in luminosity, and to cope with the demanding operational conditions at the HL-LHC [95, 146, 154, 156, 157]. The upgrade of the first level hardware trigger (L1) will allow for an increase of the L1 rate and latency to about 750 kHz and 12.5 μs , respectively, and the high-level software trigger (HLT) is expected to reduce the rate by about a factor of 100 to 7.5 kHz. The entire pixel and strip tracking detectors will be replaced to increase the granularity, reduce the material budget in the tracking volume, improve the radiation hardness, and extend the geometrical coverage and provide efficient tracking up to pseudorapidities of about $|\eta| = 4$. The muon system will be enhanced by upgrading the electronics of the existing cathode strip chambers (CSC), resistive plate chambers (RPC) and drift tubes (DT). New muon detectors based on improved RPC and gas electron multiplier (GEM) technologies will be installed to add redundancy, increase the geometrical coverage up to about $|\eta| = 2.8$, and improve the trigger and reconstruction performance in the forward region. The barrel electromagnetic calorimeter (ECAL) will feature the upgraded front-end electronics that will be able to exploit the information from single crystals at the L1 trigger level, to accommodate trigger latency and bandwidth requirements, and to provide 160 MHz sampling allowing high precision timing capability for photons. The hadron calorimeter (HCAL), consisting in the barrel region of brass absorber plates and plastic scintillator layers, will be read out by silicon photomultipliers (SiPMs). The endcap electromagnetic and hadron calorimeters will be replaced with a new combined sampling calorimeter (HGCAL) that will provide highly-segmented spatial information in both

transverse and longitudinal directions, as well as high-precision timing information. Finally, the addition of a new timing detector for minimum ionizing particles (MTD) in both barrel and endcap regions is envisaged to provide the capability for 4-dimensional reconstruction of interaction vertices that will significantly offset the CMS performance degradation due to high pile-up rates.

6.3 Extrapolation of the Run 2 $H \rightarrow \mu\mu$ results to the HL-LHC

The projections of the Run 2 $H \rightarrow \mu\mu$ results to the HL-LHC are performed without changing the analysis strategy described in Section 4.6, but only by manipulating the shapes and normalization of signal and background models, and adjusting the systematic uncertainties. Only the two most sensitive channels — ggH and VBF — are considered for this extrapolation, as the impact of the other channels on the expected analysis sensitivity is small.

The changes in the signal and background event yields from Run 2 to HL-LHC are predominantly caused by two effects: an increase in cross section of the processes due to the increased center-of-mass energy, and the extended muon acceptance of the Phase-2 detector. The considered shape effects are related to the improvements in the dimuon mass resolution. The two normalization effects are discussed in detail in Sections 6.3.1 and 6.3.2, while the updated dimuon mass resolution is described in Section 6.3.3. The considered effects are not evaluated separately for the ggH and VBF event categories, due to the fact that the categorization is very sensitive to the jet kinematics, which are modeled imperfectly in the simulated HL-LHC datasets.

The total effect on signal and background normalization is evaluated by comparing the event yields after full event selection between the simulated Run 2 datasets produced using full CMS detector simulation based on GEANT4 [116, 117], and the simulated HL-LHC datasets produced using DELPHES [158], fast parametric detector simulation package. For Run 2, only the simulated datasets produced for the 2018 data-taking period are used, in order to better reflect the detector conditions at the end of Run 2.

The event selection for the DELPHES datasets closely follows that for Run 2, with a few changes summarized in Table 6.2.

In the Run 2 selection for the 2018 datasets, events were required to contain a pair of isolated oppositely-charged muons within $|\eta| < 2.4$ (detector geometry limitation), such that at least one muon has transverse momentum exceeding 26 GeV (efficiency plateau of HLT_IsoMu24). The categorization of events between the channels targeting different Higgs production modes was based on the kinematics of jets present in the event (e.g. for the VBF channel, the events were required to contain two jets with invariant mass greater than 400 GeV, angular separation $|\Delta\eta| > 2.5$, and $p_T > 35$ GeV for the leading jet).

In the HL-LHC data-taking period, in addition to the increased $\sqrt{s}$, the detector conditions will change as well, due to the improvements provided by the Phase-2 upgrade [95, 154, 157] of the CMS detector. The updated detector geometry will allow to select muons in the forward region up to $|\eta| = 2.8$, as opposed to the $|\eta| < 2.4$ limitation in Run 2, resulting in a considerable increase of the event yields. The muon HLT performance at HL-LHC is assumed to remain the same as in Run 2 [6], and the performance is assumed not to be affected by the increased η coverage. Although the DELPHES datasets do not contain HLT objects, the assumption of the unchanged trigger conditions is implemented by keeping the leading muon selection at 26 GeV.

Additionally, the pseudorapidity cuts are extended for jets and electrons (the latter are only used to veto the events containing electrons), and the jet p_T selection is increased from 25 GeV to 30 GeV, as recommended by the upgrade performance group [159]. The updated selections of jets and electrons do not significantly affect the results of the extrapolation.

The resulting event yields for the Run 2 MC (2018) and DELPHES HL-LHC datasets, both scaled to $L = 60 \text{ fb}^{-1}$, are given in Table 6.3.

Only the most important backgrounds are considered:

- the Drell–Yan process, which accounts for $\approx 90\%$ of the total background;
- the top quark-antiquark pair production (TTJets), which accounts for $\approx 10\%$ of the background in the ggH category;

Table 6.2. Event and object selections in Run 2 and HL-LHC

Variable	Run 2 selections	HL-LHC selections
Muon ID working point	Medium	Medium
Muon $ \eta $	≤ 2.4	≤ 2.8
Muon p_T	≥ 20 GeV	≥ 20 GeV
Leading muon p_T	≥ 26 GeV (29 GeV in 2017)	≥ 26 GeV
Muon relative isolation	≤ 0.25	≤ 0.25
Jet ID working point	Loose (2016), Tight (2017/18)	Tight
Jet $ \eta $	≤ 4.7	≤ 5
Jet p_T	≥ 25 GeV	≥ 30 GeV
Pile-up suppression	Loose PUID	PUPPI
Jet cleaned w.r.t muons	$\Delta R < 0.4$	$\Delta R < 0.4$
Electron ID working point	WP90	Tight
Electron $ \eta $	≤ 2.5	≤ 3
Electron p_T	≥ 25 GeV	≥ 25 GeV
HLT path	IsoMu24, IsoTkMu24, IsoMu27	No changes assumed
Good primary vertex	Required	Not used
Event flags	BadPFMuonFilter, etc.	Not used

- the electroweak production of a Z boson (EWK_ZToLL), which is the leading background in the sensitive regions of the phase space in the VBF category.

It has been observed that the changes in the event yields for other backgrounds (such as diboson and single top processes) have a negligible effect on the analysis sensitivity: increasing the event yields of these processes by 100% affects the expected significance by less than 0.1%. Therefore, in this extrapolation exercise, only the event yield changes for the three main backgrounds are considered.

Table 6.3. Total change in event yields between Run 2 and HL-LHC analyses. The yields are quoted for $L = 60 \text{ fb}^{-1}$.

Process		Run 2 yield	HL-LHC yield	Change
Backgrounds	Drell–Yan (MG)	720623	900825	+25.0%
	TTJets Dilepton	62585	87592	+40.0%
	EWK $Z \rightarrow \ell\ell$	1439	1865	+29.6%
	Total background	784647	990282	+26.2%
$H \rightarrow \mu\mu$ signal	ggH	383	477	+24.5%
	VBF	30.7	37.8	+23.1%

The effects on event yields described above are computed without identifying the specific reasons for the changes.

6.3.1 Scaling of cross sections

During Run 2, the LHC operated at a center-of-mass energy of 13 TeV; however, for the HL-LHC data-taking period the energy will be increased to 14 TeV. Hence, the cross sections of the signal and background processes will increase, resulting in higher event yields for a given integrated luminosity (see Figure 1.3). The increase due to cross section scaling for the signal processes is expected to be $\approx 12\%$, while for the main background (Drell–Yan) it is slightly lower, $\approx 8\%$ (see Table 6.4).

Both Drell–Yan and signal cross sections are calculated at NNLO in perturbative QCD, including electroweak effects at NLO. The Drell–Yan cross sections are computed in the mass window $110 < m_{\mu\mu} < 150$ GeV using the FEWZ 3.1.b2 program [135] with the LUXqed_plus_PDF4LHC15_nnlo_100 [160] PDF set. The Higgs production cross sections are provided by the LHC Higgs Cross Section Working Group [43]. The cross sections for $TT\text{Jets}$ and $EWK_Z\text{ToLL}$ are obtained from the available generator information (both datasets were generated with the MADGRAPH [112] and PYTHIA8 [114] generators).

Table 6.4. Cross sections at 13 TeV and 14 TeV.

Process		13 TeV xSec, pb	14 TeV xSec, pb	Change
Backgrounds	Drell–Yan M110-150	26.31	28.37	+7.8%
	$TT\text{Jets}$ Dilepton	85.66	103.35	+20.7%
	EWK Z+jets M-50	3.998	4.380	+9.6%
$H \rightarrow \mu\mu$ signal	ggH	0.01057	0.01187424	+12.3%
	VBF	0.0008210722	0.00092442	+12.6%

6.3.2 Detector acceptance

The normalization effects unrelated to updated cross sections are estimated by implementing the event selections as described in Table 6.2, and then scaling the event weights in the DELPHES datasets down by the factors corresponding to the cross section scaling described in Section 6.3.1. The resulting event yields are given in Table 6.5.

Table 6.5. Effect of updated detector acceptance: total event yields for Run 2 and HL-LHC analyses with all cross sections scaled to their 13 TeV values. The yields are quoted for $L = 60 \text{ fb}^{-1}$.

Process		Run 2 yield	HL-LHC yield	Change
Backgrounds	Drell–Yan (MG)	720623	835645	+16.0%
	TTJets Dilepton	62585	72570	+16.0%
	EWK $Z \rightarrow \ell\ell$	1439	1702	+18.3%
$H \rightarrow \mu\mu$ signal	ggH	383	424	+10.7%
	VBF	30.7	33.6	+9.4%

The reasons behind the remaining yield differences can be explained by comparing the distributions of various kinematic parameters. It has been found that the largest contribution to these differences is caused by the extended pseudorapidity acceptance of the muons (the $|\eta|$ cut was updated from 2.4 to 2.8). Limiting the pseudirapidity range to $|\eta| < 2.4$ for both Run 2 and DELPHES datasets, in addition to fixing the cross sections to 13 TeV values, leads to very close shapes of the distributions, as shown in Figure 6.1. Therefore, the normalization effects other than the cross section scaling are referred to as “acceptance effects”. The remaining $\approx 3\%$ yield difference can be attributed to small changes in muon kinematics due to an additional boost associated with the increased LHC center-of-mass energy.

Table 6.6 summarizes the effects of both cross section and acceptance changes, in order to illustrate the relative contributions of each effect to the total yield increase.

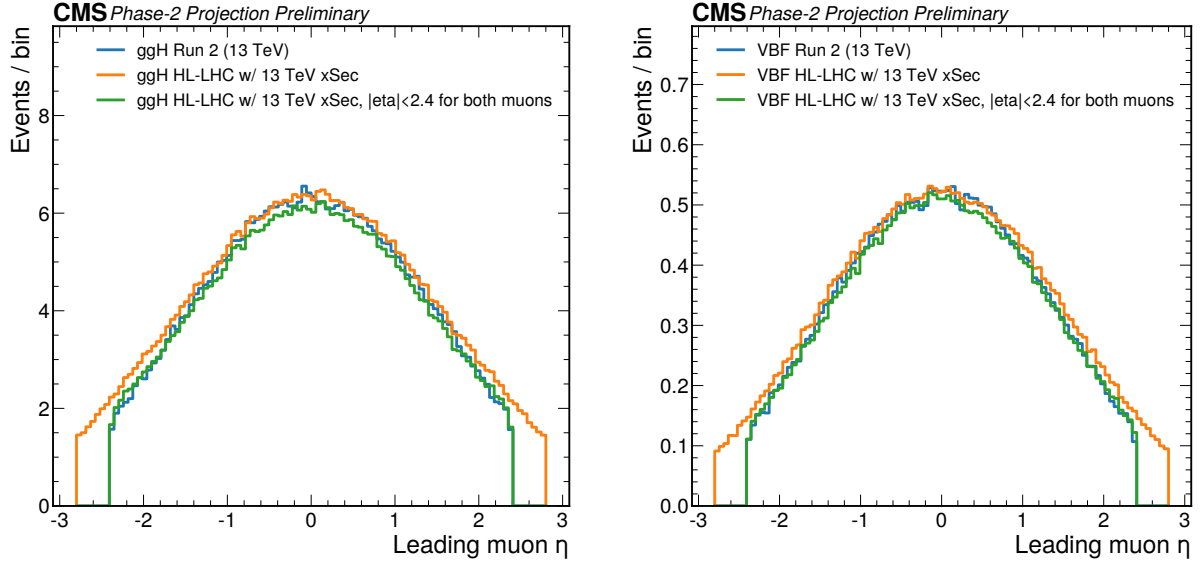


Figure 6.1. Pseudorapidity of the leading muon for ggH $H \rightarrow \mu\mu$ (left) and VBF $H \rightarrow \mu\mu$ (right) datasets after full event selection.

Table 6.6. Relative change of the inclusive event yields due to changes to cross sections and detector acceptance.

Process		Cross sections	Acceptance	Total yield change
Backgrounds	Drell–Yan (MG)	+7.8%	+16.0%	+25.0%
	TTJets Dilepton	+20.7%	+16.0%	+40.0%
	EWK $Z \rightarrow \ell\ell$	+9.6%	+18.3%	+29.6%
	Total background	+8.8%	+16.0%	+26.2%
$H \rightarrow \mu\mu$ signal	ggH	+12.3%	+10.7%	+24.5%
	VBF	+12.6%	+9.4%	+23.1%

6.3.3 Dimuon mass resolution

In the ggH, $t\bar{t}H$, WH, and ZH analysis categories, the signal and background shapes are given by the dimuon invariant mass distributions, smoothly falling for the backgrounds and forming a peak at $M = 125$ GeV for the signal processes. The shape of the signal peak is completely dominated by the dimuon mass resolution of the CMS detector. After the Phase-2 CMS tracker upgrade [154], the mass resolution is expected to improve by approximately 30% in the phase space of the $H \rightarrow \mu\mu$ signal. Figure 6.2 shows the mass resolution in 2018

MC and DELPHES datasets for the ggH $H \rightarrow \mu\mu$ signal scaled to $L = 60 \text{ fb}^{-1}$ inclusively for all selected events (left), and as a function of the max $|\eta|$ of the two muons (right).

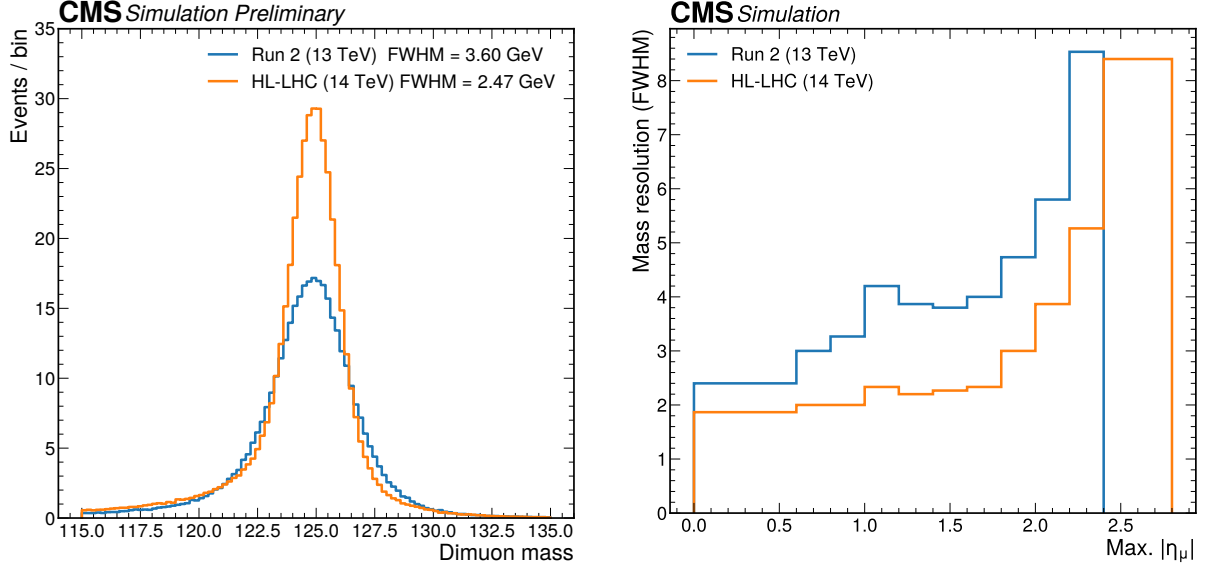


Figure 6.2. Left: Inclusive ggH $H \rightarrow \mu\mu$ dimuon mass spectrum in Run 2 and HL-LHC. Right: Full width at half maximum (FWHM) of the Higgs peak as a function of the largest $|\eta|$ of the two muons.

ggH category

The effect of the improved mass resolution has been taken into account in the ggH category by manually narrowing the shapes of the signal models. The width parameter of the double-sided Crystal Ball function [133] describing the signal models in the subcategories of the ggH channel has been decreased by a factor of 0.685 (change in FWHM of the Higgs peak evaluated inclusively for all selected signal events).

VBF category

Since the dimuon mass was used in the VBF category in a non-trivial way (as an input feature to the DNN), it is not possible to exactly estimate the effect of the improved mass resolution without retraining the DNN. Instead, an approximation approach has been devel-

oped. In this approach, the signal events are reweighted using the ratio of the VBF signal mass spectra at HL-LHC and Run 2, such that the events around 125 GeV are assigned higher weights, events in the tails of the mass spectrum are assigned lower weights, and the total event yield is left unchanged. Figure 6.3 shows the normalized mass spectra and their ratio in 0.5 GeV bins, which was used as a look-up table for event reweighting. The elevated shapes in the sidebands of the ratio plot are caused by the fact that muon corrections (final state radiation recovery, Rochester scale correction, GeoFit correction) are applied for the Run 2 datasets, but not applied for HL-LHC datasets. However, the number of events in the sidebands is small and does not significantly affect the procedure.

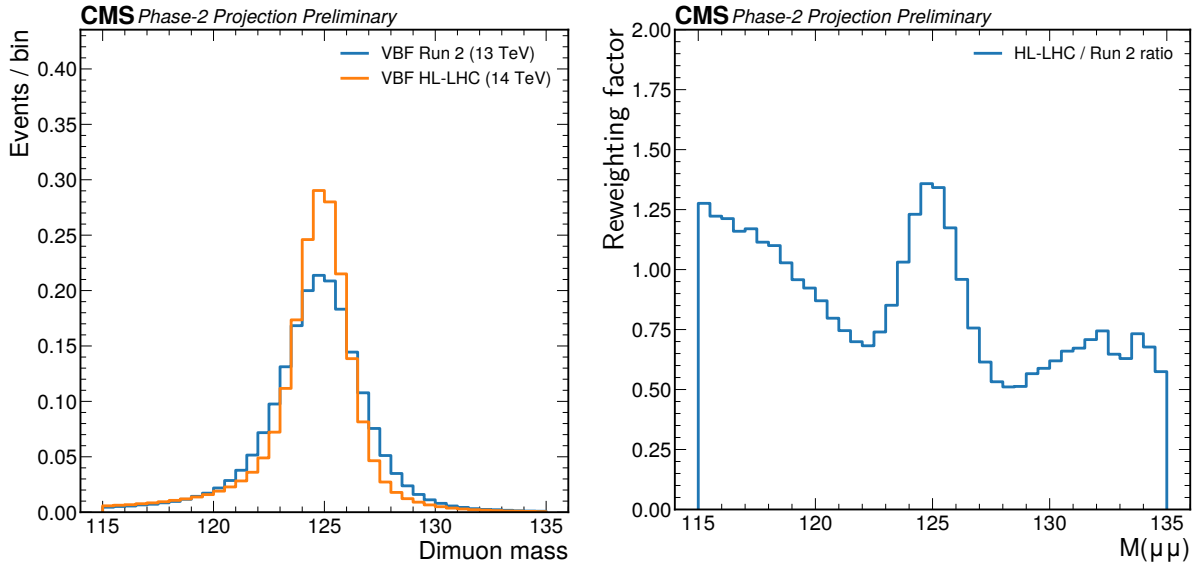


Figure 6.3. Left: Dimuon mass spectrum of the VBF $H \rightarrow \mu\mu$ signal in Run 2 and at HL-LHC. Right: Reweighting factors for estimation of the effect of mass resolution improvement in the VBF channel.

The weights computed with this approach are used to reweight the DNN score distribution and obtain the new signal templates, as shown in Figure 6.4. The reweighting procedure provides a recipe to produce only the nominal templates; the systematic variations of the signal shapes are constructed in such a way that the ratio of the modified and nominal shape is the same for the reweighted templates as it was for the original ones. An example of the modified shapes is shown in Figure 6.5.

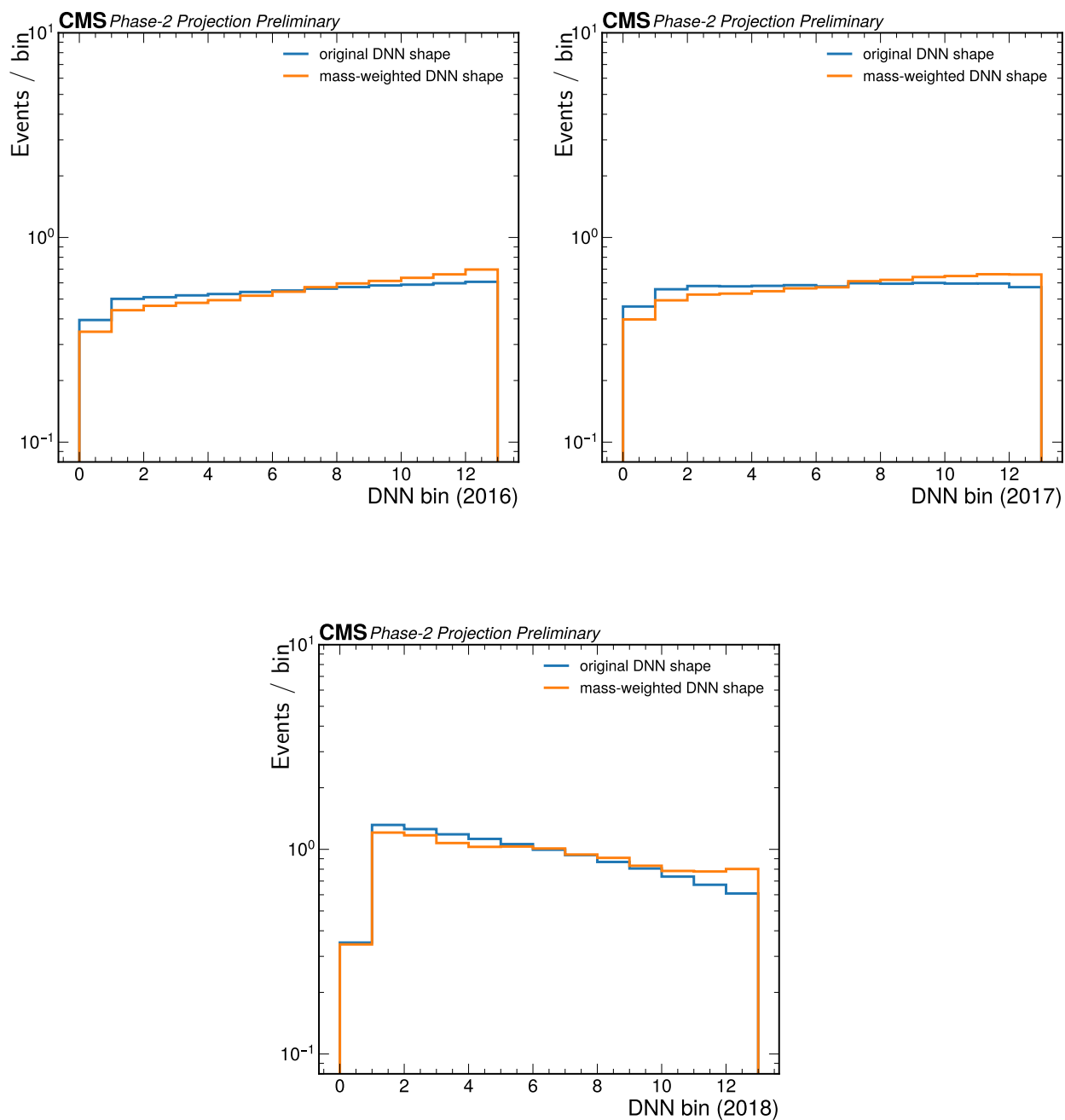


Figure 6.4. Original and reweighted according to the updated mass resolution DNN templates in VBF category for 2016 (top left), 2017 (top right) and 2018 (bottom) data-taking periods.

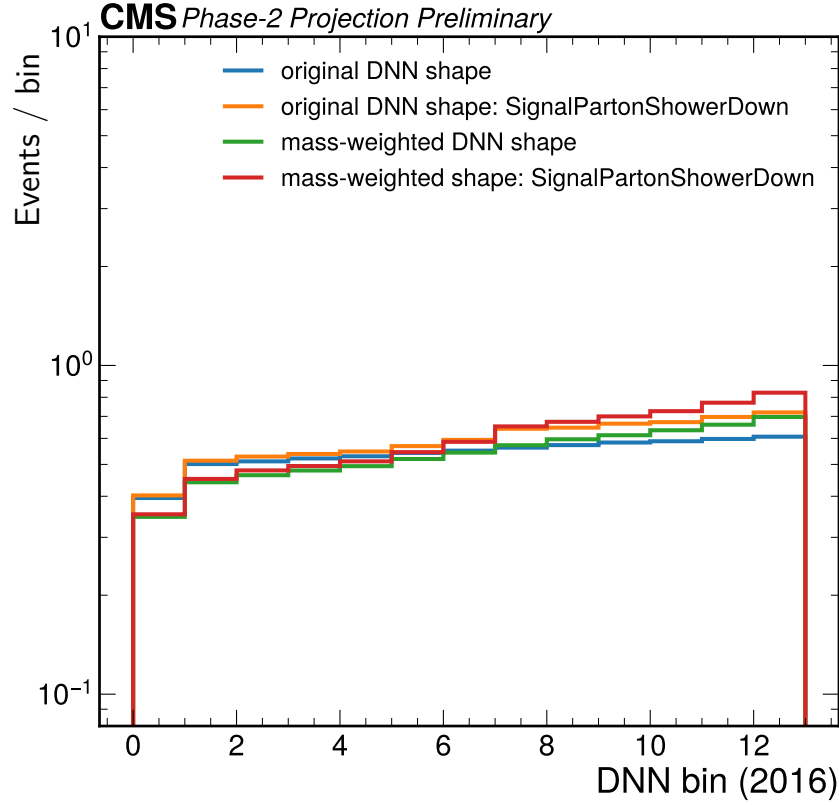


Figure 6.5. Nominal and varied DNN shapes in the original Run 2 datacards (blue and orange), and after mass-reweighting (green and red).

6.3.4 Uncertainty scenarios

The large amount of data set to be delivered by the HL-LHC will enable the measurements of very rare phenomena with unprecedented accuracy and sensitivity. However, the HL-LHC will also pose new challenges, such as a higher amount of pile-up (≈ 200 primary vertices per collision event). In order to estimate the future physics potential of the upgraded CMS detector at the HL-LHC, the projections are made under different systematic uncertainty scenarios, both conservative and optimistic. The baseline, or the conservative scenario assumes that the CMS upgrade will be able to overcome the HL-LHC challenges and provide the same level of performance as in Run 2, while the optimistic scenario assumes that the theoretical and experimental uncertainties will be further decreased by the future research efforts.

For the projections of physics analyses to the HL-LHC conditions, two scenarios are considered [153]. In the first scenario (S1), all systematic uncertainties are kept at the same values as in Run 2. In the second scenario (S2), all the systematic uncertainties are scaled down proportionally to the square root of the integrated luminosity until a given lower limit is reached, while the theoretical uncertainties are halved from their Run 2 values. The lower limits for the systematic uncertainties are estimated from the expected accuracy of the upgraded detector.

In S2, the systematic uncertainties on the electron and muon trigger and identification efficiencies are expected to be reduced to 0.5%, and are fully correlated between the leptons of the same flavour. About 20–50% reduction in the jet energy scale (JES) uncertainty is expected due to the significant reduction of the pile-up uncertainty. Additional improvements for the uncertainties in the forward detector region will come from the HGCal and wider tracker coverage. The uncertainty in the b -jet tagging efficiency is predicted to be 1% for all operating points and p_T . The uncertainty on the integrated luminosity is assumed to be as low as 1% due to a better understanding of the detector calibration procedure, more granular detector and more radiation tolerant electronics.

The uncertainty on the background model choice in the ggH category can not be scaled with the integrated luminosity as it is done for most of the other systematic uncertainties, as it stems from the choice of different probability distribution functions and is treated as a discrete nuisance in the likelihood. In both S1 and S2, the background model is not allowed to vary, therefore no uncertainty due to background modeling is taken into account. Instead, the sensitivity of the ggH category was estimated for each of the considered background models, and the variation in the expected significance was found to be within 5%. All the uncertainties used in this projection study are summarized in Table 6.7.

Table 6.7. Comparison between different uncertainty scaling scenarios.

Source of uncertainty	Scenario 1 (Run 2 syst.)	Scenario 2 (YR18 syst.)
Jet energy scale	5–6%	$1/5$ – $1/2 \times S1$
Jet energy resolution	2–3%	3–5%
Muon ID, isolation, trigger	0.1–7.0%	0.5%
electron ID, isolation, trigger	0.1–0.5%	0.5%
b -tagging efficiency	5%	1%
L1 prefiring correction	20%	4%
Theory	9.4%	$1/2 \times S1$
Integrated luminosity	2.5%	1%
Limited number of MC events	Barlow-Beeston method [125]	–

6.4 Results

The expected significance, as well as the expected measurement precision for two Higgs boson properties, namely the signal strength modifier $r^{\mu\mu}$ and the coupling modifier for the Higgs coupling to muons κ_μ , are estimated for the HL-LHC conditions — $L = 3000 \text{ fb}^{-1}$ at $\sqrt{s} = 14 \text{ TeV}$. The results are compared with the previous projections — the extrapolation performed after the Run 1 (Snowmass 2013 [148,149]), and the extrapolation performed after the 2016 data-taking period of the Run 2 (LHC Yellow Report 2018 [153]).

The total relative uncertainty on the signal strength as a function of the integrated luminosity is shown in Figure 6.6 on the left, while the projection for the κ_μ uncertainty is shown on the right. The breakdown of the uncertainties by source, as well as the comparison with the previous projections, is shown in Tables 6.8 and 6.9. The extrapolated uncertainties are improved by 30–35% with respect to the Yellow Report 2018 projections, due to the improved analysis strategy implemented during Run 2, such as an advanced event categorization and improved machine learning methods.

The largest impact on the analysis sensitivity improvement at the HL-LHC is expected to be due to the improved dimuon mass resolution after the Phase-2 upgrade. This effect is illustrated in Figure 6.7, where the uncertainties on $r^{\mu\mu}$ and κ_μ are shown for the ggH category before and after modifying the signal models to take into account the mass resolution improvement. In both plots, the measurements are shown for the S2 uncertainty scenario. A

similar comparison for the expected significance of the ggH channel is shown in Figure 6.8 on the right. At $L = 3000 \text{ fb}^{-1}$, the relative improvement due to the improved mass resolution in the ggH channel is approximately 25% for all considered parameters of interest.

The extrapolation of the total expected significance as a function of integrated luminosity is shown in Figure 6.8 on the left. The significance is computed using an Asimov dataset and assuming that the $H \rightarrow \mu\mu$ signal strength is equal to the SM expectation. With the projected analysis sensitivity, the observation-level significance of 5σ can be achieved with $L \approx 300 \text{ fb}^{-1}$ of data collected at $\sqrt{s} = 14\text{TeV}$. However, in combination with the existing Run 1 and Run 2 measurements, and with further improvements to the analysis strategy, this level of sensitivity will likely be achieved with even smaller amount of new data.

Table 6.8. Uncertainty on the signal strength modifier $r^{\mu\mu}$ at $L = 3000 \text{ fb}^{-1}$.

		Statistical	Experimental	Theoretical	Total
S1	Snowmass 2013	—	—	—	20.0%
	YR 2018	9.1%	7.6%	4.8%	12.8%
	Snowmass 2021	6.4%	3.7%	4.2%	8.5%
S2	Snowmass 2013	—	—	—	14.0%
	YR 2018	9.1%	1.7%	2.7%	9.6%
	Snowmass 2021	6.4%	2.0%	2.0%	7.0%

Table 6.9. Uncertainty on the coupling modifier κ_μ at $L = 3000 \text{ fb}^{-1}$.

		Statistical	Experimental	Theoretical	Total
S1	Snowmass 2013	—	—	—	8.0%
	YR 2018	4.7%	2.7%	3.9%	6.7%
	Snowmass 2021	3.2%	1.9%	2.2%	4.3%
S2	Snowmass 2013	—	—	—	7.5%
	YR 2018	4.7%	1.5%	1.1%	5.0%
	Snowmass 2021	3.2%	1.1%	0.8%	3.5%

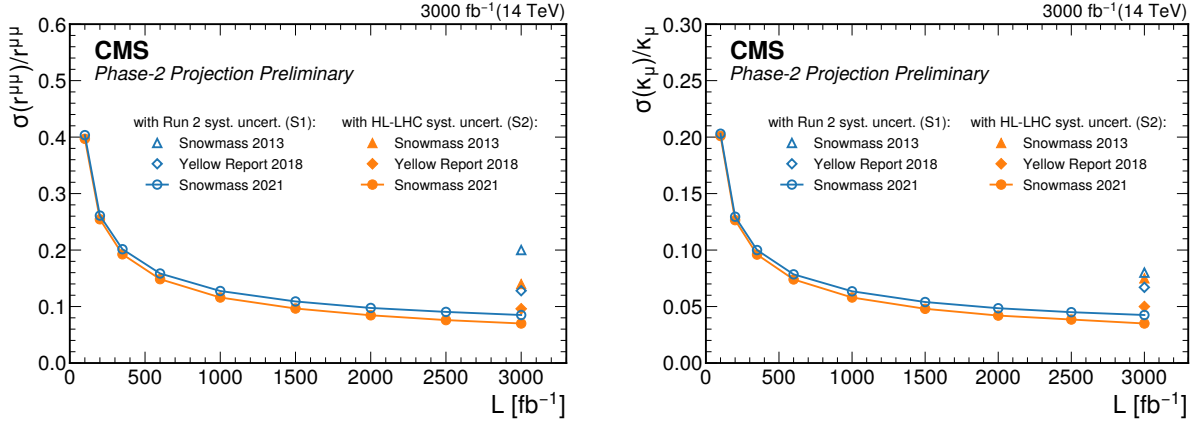


Figure 6.6. Uncertainty on signal strength modifier $r^{\mu\mu}$ (left) and κ_μ (right).

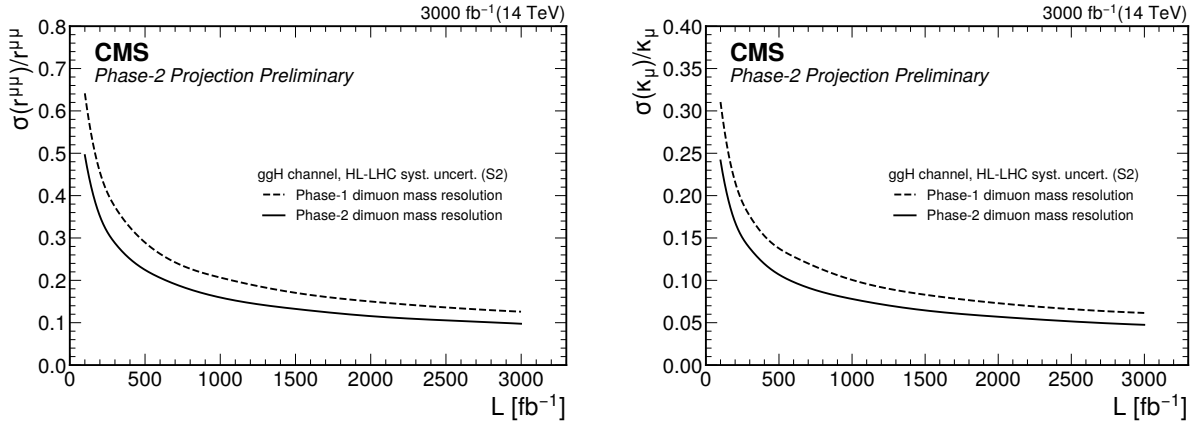


Figure 6.7. Effect of improved dimuon mass resolution in ggH channel for S2 uncertainty scenario. Left: Uncertainty on signal strength modifier $r^{\mu\mu}$. Right: Uncertainty on coupling modifier κ_μ .

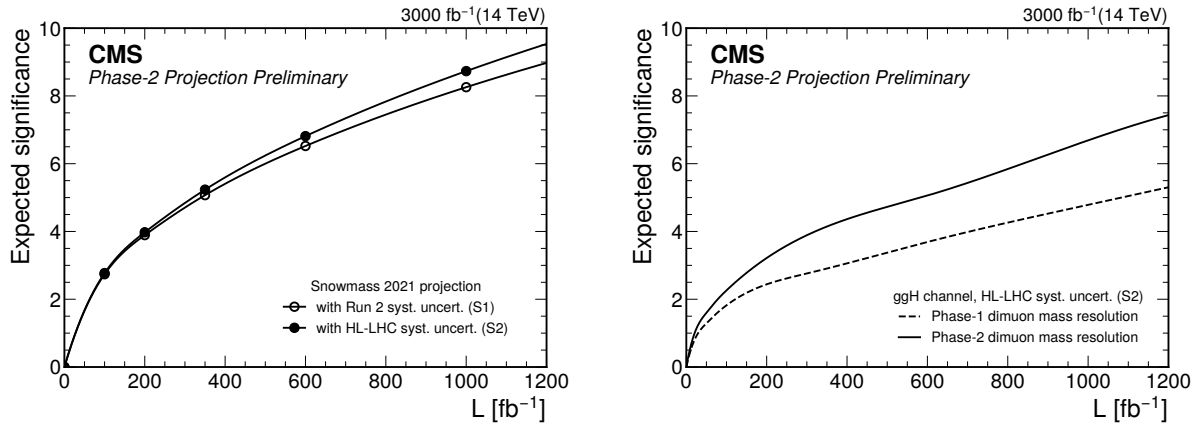


Figure 6.8. Left: Extrapolation of the expected significance for S1 and S2 uncertainty scenarios. Right: Effect of the improved dimuon mass resolution on expected significance in ggH channel for S2 uncertainty scenario.

SUMMARY

The analysis described in this thesis presents the first evidence for the Higgs boson decay to second generation fermions, specifically to a pair of muons [1]. The search is performed using the proton-proton collision data corresponding to an integrated luminosity of 137 fb^{-1} , produced at center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$, and recorded by the CMS experiment during Run 2 of the LHC (2016–2018). The observed results are in agreement with the Standard Model prediction, and they provide the best constraint on the coupling between the Higgs boson and the muon to date.

This analysis largely improves on the previous $H \rightarrow \mu\mu$ measurement published by CMS [132], which was obtained using only the 2016 data corresponding to an integrated luminosity of 35.9 fb^{-1} . The improvements are due to both the larger amount of data analyzed, and an improved analysis strategy. The previous measurement targeted mainly the ggH production mode, and failed to exploit the characteristic properties of other production channels. In contrast, the Run 2 search features dedicated analysis strategies for different Higgs production channels, leading to a significant increase in the overall analysis sensitivity.

The improvement to the $H \rightarrow \mu\mu$ analysis strategy is driven by a conceptually different approach used in the category targeting the VBF production mode. The main differences between this approach and the ones used in other analysis categories are the background modeling method, and the choice of the observable to extract the $H \rightarrow \mu\mu$ signal. In the VBF category, the background model is based on MC simulation, and the observable is constructed using a deep neural network, while in other categories the background model is obtained using parametric fits of the observed data in the dimuon mass sidebands, and the dimuon mass spectrum is used as an observable. As a result, the VBF channel turned out to be the most sensitive one in the Run 2 analysis, accounting for approximately half of the total signal significance.

The observed (expected) significance of the $H \rightarrow \mu\mu$ decay in the Run 2 search is 3.0 (2.5) standard deviations, which is the best result to date, but not yet sufficient to claim the discovery of the decay. The observed (expected) 95% CL upper limit on the signal strength modifier is 1.93 (0.81), while the total uncertainty is 42%, dominated by the statistical

component. The 95% CL interval on the coupling strength modifier κ_μ is $0.59 < \kappa_\mu < 1.50$ with a total uncertainty of 22%.

It will not take too long until either the $H \rightarrow \mu\mu$ decay, or a disagreement of the observations in the dimuon channel with the Standard Model is observed with a discovery-level significance of 5 standard deviations or more. The upcoming LHC data-taking periods are set to provide a sufficient amount of data to increase the sensitivity of the $H \rightarrow \mu\mu$ search to this level in the next several years. In July 2022, the LHC has started the Run 3 data-taking period, which will continue through 2025. The center-of-mass energy in Run 3 is increased to 13.6 TeV, while the integrated luminosity will be comparable to that in Run 2. After Run 3, the LHC will prepare for the High-Luminosity LHC (HL-LHC) project, which is planned to start in 2027 and aims to collect 3000 fb^{-1} of collision data at $\sqrt{s} = 14 \text{ TeV}$.

If the $H \rightarrow \mu\mu$ analysis strategy remains unchanged from that in Run 2, the 5σ level of significance will be achieved already in the early stages of the HL-LHC data-taking, and by the end of the HL-LHC program the uncertainty on the Higgs coupling to muons will be as low as 3–4% (after Run 2, this uncertainty is approximately 22%). With future improvements of the analysis methods, and by combining the results of the CMS and ATLAS collaborations, the precision of the coupling measurement can be improved even further, and the $H \rightarrow \mu\mu$ decay can be discovered even earlier, unless new experimental results emerge to challenge the Standard Model.

A. DATASETS

Table A.1. Signal simulated samples for the 2016 data taking era. The (*) stands for PUMoriond17_94X_mcRun2_asymptotic_v3.

Process	Sample	$\sigma \times \text{BR}$ (pb)
ggH	/GluGlu_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v2/MINIAODSIM	0.01057 pb
	/GluGlu_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*ext1-v2/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5_PWeights_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M120_TuneCP5_PWeights_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM	0.012653 pb
	/GluGluHToMuMu_M130_TuneCP5_PWeights_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM	0.008505 pb
	/GluGluHToMuMu_M125_TuneCP5up_PWeights_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5down_PWeights_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.01057 pb
	/VBFHTToMuMu_M-125_TuneCP5_PWeights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0008228 pb
	/VBFHTToMuMu_M125_TuneCP5_PWeights_13TeV_amcatnlo_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0008228 pb
	/VBFHTToMuMu_M120_TuneCP5_PWeights_13TeV_amcatnlo_pythia8/*-v1/MINIAODSIM	0.000953 pb
VBF	/VBFHTToMuMu_M130_TuneCP5_PWeights_13TeV_amcatnlo_pythia8/*-v1/MINIAODSIM	0.000683 pb
	/VBFHTToMuMu_M125_TuneCP5up_PWeights_13TeV_amcatnlo_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0008228 pb
	/VBFHTToMuMu_M125_TuneCP5down_PWeights_13TeV_amcatnlo_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0008228 pb
	/VBFHTToMuMu_M-125_TuneEEC5_13TeV-amcatnlo-herwigpp/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0008228 pb
	/VBFHTToMuMu_M-125_TuneEEC5_13TeV-powheg-herwigpp/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0008228 pb
	/WPlusH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v2/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M120_TuneCP5_PWeights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000232 pb
	/WplusH_HToMuMu_WToAll_M130_TuneCP5_PWeights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000139 pb
	/WplusH_HToMuMu_WToAll_M125_TuneCP5up_PWeights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M125_TuneCP5down_PWeights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001858 pb
WH	/WminusH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v2/MINIAODSIM	0.0001164 pb
	/WminusH_HToMuMu_WToAll_M120_TuneCP5_PWeights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000148 pb
	/WminusH_HToMuMu_WToAll_M130_TuneCP5_PWeights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000088 pb
	/WminusH_HToMuMu_WToAll_M125_TuneCP5up_PWeights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001164 pb
	/WminusH_HToMuMu_WToAll_M125_TuneCP5down_PWeights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001164 pb
	/ZH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v2/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M120_TuneCP5_PWeights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000241 pb
	/ZH_HToMuMu_ZToAll_M130_TuneCP5_PWeights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000148 pb
	/ZH_HToMuMu_ZToAll_M125_TuneCP5up_PWeights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M125_TuneCP5down_PWeights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001923 pb
ZH	/ttHTToMuMu_M125_TuneCP5_PWeights_13TeV_powheg-pythia8/*-v1/MINIAODSIM	0.0001103 pb
	/ttHTToMuMu_M120_TuneCP5_PWeights_13TeV_powheg-pythia8/*-v1/MINIAODSIM	0.000138 pb
	/ttHTToMuMu_M130_TuneCP5_PWeights_13TeV_powheg-pythia8/*-v1/MINIAODSIM	0.000085 pb
	/ttHTToMuMu_M125_TuneCP5up_PWeights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001103 pb
ttH	/ttHTToMuMu_M125_TuneCP5down_PWeights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001103 pb

Table A.4. Background simulation samples for the 2016 data taking era. The (*) stands for RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3.

Process	Sample	Cross section (pb)	Cross section Precision
Drell-Yan	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/*_ext2-v1/MINIAODSIM	5954	NNLO (QCD)
	/DYJetsToLL_M-105To160_TuneCUETP8M1_PWeights_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	47.12	NLO (EWK)
	/DYJetsToLL_M-105To160_VBFFilter_TuneCUETP8M1_PWeights_13TeV-amcatnloFXFX-pythia8/*_ext1-v1/MINIAODSIM	1.774	NNLO (QCD)
	/EWK_LLJJ_MLL-50_MJJ-120_13TeV-madgraph-herwigpp/*_ext1-v1/MINIAODSIM	1.611	NLO (EWK)
	/EWK_LLJJ_MLL-105_160_SM_5f_L0*madgraph-herwigpp/*_ext1-v1/MINIAODSIM	0.0464	NLO (EWK)
Top	/TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8/*-v2/MINIAODSIM	86.61	NNLO (QCD)
	/TTToSemilepton_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8/*-v2/MINIAODSIM	358.57	NNLO (QCD)
	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/*_ext1-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/*_ext1-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/*-v1/MINIAODSIM	80.95	NLO (QCD)
	/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/*-v1/MINIAODSIM	136.02	NLO (QCD)
	/ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1/*-v1/MINIAODSIM	3.40	NLO (QCD)
	/tZq_l1_4f_13TeV-amcatnlo-pythia8/*_ext1-v1/MINIAODSIM	0.0758	NLO (QCD)
	/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/*_ext2-v1/MINIAODSIM	0.2001	NLO (QCD)
	/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8/*_ext2-v1/MINIAODSIM	0.2529	NLO (QCD)
	/TTTT_TuneCUETP8M2T4_13TeV-amcatnlo-pythia8/*-v1/MINIAODSIM	0.0091	NLO (QCD)
	/TTTW_TuneCUETP8M2T4_13TeV-amcatnlo-pythia8/*_ext1-v1/MINIAODSIM	0.00073	NLO (QCD)
	/TTWW_TuneCUETP8M2T4_13TeV-amcatnlo-pythia8/*_ext1-v1/MINIAODSIM	0.0070	NLO (QCD)
	/TTTT_TuneCUETP8M2T4_13TeV-amcatnlo-pythia8/*_ext1-v1/MINIAODSIM	0.000039	NLO (QCD)
Diboson	/WWTo2L2Nu_13TeV-powheg/*-v2/MINIAODSIM	12.178	NNLO (QCD)
	/WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM	4.102	NLO (QCD)
	/WZTo2L2Q_13TeV-amcatnloFXFX-madspin-pythia8/*-v2/MINIAODSIM	5.606	NLO (QCD)
	/ZZTo2L2Nu_13TeV-powheg-pythia8_ext1/*-v2/MINIAODSIM	0.5644	NLO (QCD)
	/ZZTo2L2Q_13TeV-amcatnloFXFX-madspin-pythia8/*-v3/MINIAODSIM	3.224	NLO (QCD)
	/ZZTo4L_13TeV-powheg-pythia8/*-v1/MINIAODSIM	1.256	NLO (QCD)
	/GluGluToContInToZZTo(2e2mu,2mu2tau,2e2tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.003185	LO
	/GluGluToContInToZZTo(4mu,4tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00157	LO
	/GluGluToContInToZZTo(2mu2nu)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.001493	LO
Triboson	/WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8/*-v2/MINIAODSIM	0.2086	NLO (QCD)
	/WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/*-v2/MINIAODSIM	0.1651	NLO (QCD)
	/WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/*-v2/MINIAODSIM	0.05565	NLO (QCD)
	/ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/*-v2/MINIAODSIM	0.01398	NLO (QCD)

Table A.5. Background simulation samples for the 2017 data taking era. The (*) stands for RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14.

Process	Sample	Cross section (pb)	Cross section Precision
Drell-Yan	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	6225.4	NNLO (QCD)
	/DYJetsToLL_M-105To160_TuneCP5_PWeights_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	47.12	NLO (EWK)
	/DYJetsToLL_M-105To160_VBFFilter_TuneCP5_PWeights_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	2.043	NNLO (QCD)
	/EWK_LLJJ_MLL-50_MJJ-120*madgraph-herwig7/*-v1/MINIAODSIM	1.700	NLO (EWK)
	/EWK_LLJJ_MLL-105_160_SM_5f_LO*madgraph-herwig7/*-v1/MINIAODSIM	0.0492	LO
Top	/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	831.76	NNLO (QCD)
	/TTTo2L2Nu_TuneCP5_PWeights_13TeV-powheg-pythia8/*-v1/MINIAODSIM	86.61	NNLO (QCD)
	/TTToSemiLeptonic_TuneCP5_PWeights_13TeV-powheg-pythia8/*-v2/MINIAODSIM	358.57	NNLO (QCD)
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_PWeights_13TeV-powheg-pythia8/*-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PWeights_13TeV-powheg-pythia8/*-v2/MINIAODSIM	35.9	NNLO (QCD)
	/ST_t-channel_antitop_5f_TuneCP5_PWeights_13TeV-powheg-pythia8/*-v1/MINIAODSIM	80.95	NNLO (QCD)
	/ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	136.02	NLO (QCD)
	/ST_s-channel_antitop_leptonDecays_13TeV-PWeights_powheg-pythia/*-v1/MINIAODSIM	1.293	NLO (QCD)
	/ST_s-channel_top_leptonDecays_13TeV-PWeights_powheg-pythia/*-v1/MINIAODSIM	2.069	NLO (QCD)
	/tZq_ll_4f_ckm_NLO_TuneCP5_PWeights_13TeV-amcatnlo-pythia8/*-v1/MINIAODSIM	0.0758	NLO (QCD)
	/TTWJetsToLNU_TuneCP5_PWeights_13TeV-amcatnloFXFX-madspin-pythia8/*-v1/MINIAODSIM	0.2001	NLO (QCD)
	/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8/*-v1/MINIAODSIM	0.2529	NLO (QCD)
	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/MINIAODSIM	0.0091	NLO (QCD)
	/TTTW_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/MINIAODSIM	0.00073	NLO (QCD)
	/TTTJ_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/MINIAODSIM	0.0070	NLO (QCD)
	/TTWW_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/MINIAODSIM	0.000039	NLO (QCD)
Diboson	/WWTo2L2Nu_NNPDF31_TuneCP5_PWeights_13TeV-powheg-pythia8/*-ext1-v1/MINIAODSIM	12.178	NNLO (QCD)
	/WZTo3L2Nu_13TeV-powheg-pythia8/*-v2/MINIAODSIM	4.658	NLO (QCD)
	/WZTo2L2Q_13TeV-amcatnloFXFX-madspin-pythia8/*-v1/MINIAODSIM	6.321	NLO (QCD)
	/ZZTo2L2Nu_13TeV-powheg-pythia8/*-v1/MINIAODSIM	0.601	NLO (QCD)
	/ZZTo2L2Q_13TeV-amcatnloFXFX-madspin-pythia8/*-v1/MINIAODSIM	3.696	NLO (QCD)
	/ZZTo4L_13TeV-powheg-pythia8/*-v1/MINIAODSIM	1.325	NLO (QCD)
	/GluGluToContInToZZTo(2e2mu,2mu2tau,2e2tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.003185	LO
	/GluGluToContInToZZTo(4mu,4tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00157	LO
	/GluGluToContInToZZTo(2mu2nu)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.001493	LO
Triboson	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/*-v2/MINIAODSIM	0.2086	NLO (QCD)
	/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/*-v2/MINIAODSIM	0.1651	NLO (QCD)
	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/*-v1/MINIAODSIM	0.05565	NLO (QCD)
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/*-v2/MINIAODSIM	0.01398	NLO (QCD)

Table A.6. Background simulation samples for the 2018 data taking era. The (*) stands for RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15.

Process	Sample	Cross section (pb)	Cross section Precision
Drell-Yan	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	6225.4	NNLO (QCD)
	/DYJetsToLL_M-105To160_TuneCP5_PSweights_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	47.12	NLO (EWK)
	/DYJetsToLL_M-105To160_VBFFilter_TuneCP5_PSweights_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	2.029	NNLO (QCD)
	/EWK_LLJJ_MLL-50_MJJ-120*madgraph-herwig7/*-v1/MINIAODSIM	1.700	NLO (EWK)
	/EWK_LLJJ_MLL-105_160_SM_5f_L0*madgraph-herwig7/*-v1/MINIAODSIM	0.0492	LO
Top	/TT_DiLept_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	86.61	NNLO (QCD)
	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM		NNLO (QCD)
	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	358.57	NNLO (QCD)
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_t-channel_antitop_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	80.95	NLO (QCD)
	/ST_s-channel_top_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	136.02	NLO (QCD)
	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8/*_ext1-v4/MINIAODSIM	3.40	NLO (QCD)
	/tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-madgraph-pythia8/*_ext1-v2/MINIAODSIM	0.0758	NLO (QCD)
	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/*_ext1-v2/MINIAODSIM	0.2001	NLO (QCD)
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.2529	NLO (QCD)
	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.0091	NLO (QCD)
	/TTTW_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.00073	NLO (QCD)
	/TTWW_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.00698	NLO (QCD)
	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.00039	NLO (QCD)
Diboson	/WWTo2L2Nu_NNPFD31_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	12.178	NNLO (QCD)
	/WZTo3LNu_TuneCP5_13TeV-powheg-pythia8/*_ext1-v2/MINIAODSIM	4.658	NLO (QCD)
	/WZTo2L2Q_13TeV-amcatnloFXFX_madspin_pythia8/*-v1/MINIAODSIM	5.606	NLO (QCD)
	/ZZTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/*_ext1-v2/MINIAODSIM	0.5644	NLO (QCD)
	/ZZTo2L2Q_13TeV-amcatnloFXFX_madspin_pythia8/*-v1/MINIAODSIM	3.659	NLO (QCD)
	/ZZTo4L_TuneCP5_13TeV-powheg-pythia8/*_ext1-v2/MINIAODSIM	1.256	NLO (QCD)
	/GluGluToContInToZZTo(2e2mu,2mu2tau,2e2tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.003185	LO
	/GluGluToContInToZZTo(4mu,4tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00157	LO
Triboson	/GluGluToContInToZZTo(2mu2nu)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.001493	LO
	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.2086	NLO (QCD)
	/WWZ_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.1651	NLO (QCD)
	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.05565	NLO (QCD)
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.01398	NLO (QCD)

B. MACHINE LEARNING METHODS IN PARTICLE PHYSICS

High energy physics (HEP) studies processes of a probabilistic nature, many of which are quite rare, and therefore require to analyze large datasets with advanced statistical methods. Therefore, adoption of state-of-the-art techniques is necessary to extract as much useful information as possible from the available data. In recent years, the use of machine learning methods has become widespread in the HEP community. In particular, some of the methods used in the CMS collaboration include boosted decision trees (BDT) [161], deep neural networks (DNN) [162–164], as well as convolutional neural networks (CNN), graph neural networks (GNN), and numerous other methods [165].

In the scope of this thesis, the two methods used for the studies were DNN and BDT. The details and applications of these methods are discussed in this chapter.

B.1 Deep Neural Networks (DNN)

A deep neural network, similarly to other machine learning tools, is essentially an analytic function that maps a multidimensional space of input parameters to one or multiple output values in such a way that the output values closely approximate given target (“truth”) values for each set of input values. The specific structure of such a function in case of DNN is represented with a network containing a large number of nodes (*neurons*) organized in multiple layers, such that a value assigned to each node is a linear combination of the values of the nodes in the previous layer, transformed using a non-linear *activation function*. The coefficients in the linear combination are called *weights*, and their values are obtained using an iterative *training* procedure.

The training procedure starts with randomly initializing all weights in the network. In order to evaluate the correctness of the network output, a *loss function* is defined, which takes the DNN output values (*predictions*) and the truth values as arguments. The loss function is typically chosen such that it is differentiable with respect to the network inputs, and the value of the loss function (called simply *loss*) is minimized when the DNN predictions

exactly match the truth values. During the training, the *backpropagation* approach is used, in which the gradients of the loss function with respect to the values in different nodes are computed, and the corresponding weights are updated based on these gradients in order to minimize the loss. The methods used to minimize the loss are typically based on *gradient descent*, where the updated value of a weight is obtained by taking a step in the direction opposite to the corresponding gradient. This procedure is repeated until the loss converges to a certain value which is assumed to be the minimum of the loss, at which point the values of weights are fixed, and the network can be evaluated on new data.

The dataset used for training is required to be different from the dataset on which the DNN score is eventually calculated, in order to prevent *overfitting* – a bias caused by the DNN “memorizing” the features of the training dataset rather than learning the correlations in data.

In the studies described in this thesis, deep neural networks are used in two different ways. In the first approach, DNN is used to reduce the dimensionality of data, in such a way that a one-dimensional observable produced by the DNN preserves as much discriminating power of the input features as possible. This approach is discussed in Section B.1.1. In the second approach, DNN is used as a tool for exploratory research, in order to find and utilize the correlations between a set of input features and an arbitrary set of parameters of interest. This approach is discussed in Section B.1.2.

B.1.1 DNN as a method to approximate a likelihood ratio

According to the Neyman–Pearson lemma [166], the most powerful statistical test for two hypotheses is given by the *likelihood ratio* test statistics

$$\lambda = \frac{P(\bar{x}|H_1)}{P(\bar{x}|H_0)},$$

where $\bar{x}$ is the observed data, H_0 and H_1 are the null and the alternative hypotheses, respectively, and P denotes probability. Although in particle physics, in principle, $P(\bar{x}|H_i)$ are well defined by the theory, in practice their computation is rarely possible. Instead, the approximate probability distributions are obtained using Monte Carlo simulation (MC).

In HEP analyses like $H \rightarrow \mu\mu$, the dimensionality of $\bar{x}$ can be high (e.g. directions and momenta of all decay products), and the amount of data in the sensitive regions insufficient to perform a multidimensional likelihood-ratio test explicitly. Therefore, it is a common practice to reduce the dimensionality by applying a function that maps $\bar{x}$ to a scalar value.

An example of such function is a deep neural network with the number of input nodes equal to dimensionality of $\bar{x}$, and a single output node yielding a scalar *DNN score* for each input. It can be shown [167] that a DNN trained by minimizing a *binary cross-entropy* (BCE)¹ loss function or *mean squared error* (MSE) loss function produces an output score which is monotonically related to the likelihood ratio distribution.

The considerations described above suggest that obtaining the most powerful test statistics should be a straightforward task. However, there are multiple assumptions to be made and limitations to be taken into account:

- The MC modeling of signal and background PDFs may be affected by systematic biases due to imprecise knowledge of the theory, therefore it may not always correctly represent the true probability distributions.
- The MC distributions are also subject to statistical and systematic uncertainties, which limit their statistical power as compared to “true” PDFs.
- Although the trained DNN is supposed to yield an output with statistical power similar to that of the likelihood ratio, in practice its performance depends on many factors, such as the DNN architecture, amount of training data, and the choice of the optimizer to find the global minimum of the loss function.

The DNN application described here was used to produce a sensitive observable in the MC template-based strategy of the Run 2 $H \rightarrow \mu\mu$ search (Section 4.6) and the related studies (Section 5.2).

¹↑ BCE and MSE loss functions are defined in Section B.1.4

B.1.2 DNN as a tool to exploit correlations between parameter spaces

Another application of neural networks considered in this thesis is automatic identification of useful correlations in data, and using them for optimization of the analysis methods.

In many cases, an analysis method can be described with a set of discrete or continuous parameters, such as the number of seeds in the muon HLT reconstruction (Section 3.4), or indices of a pair of jets in the VBF selection in the $H \rightarrow \mu\mu$ search (Section 5.1.2). The optimal configuration of the parameters for a given method may be different depending on the event properties such as particle momenta and pseudorapidities. The goal of the DNN-based optimization approach is to construct a model which yields an optimal configuration of the method parameters for a given set of event properties. If the parameters of interest are discrete, the DNN can be used as a classifier; if the parameters of interest are continuous, the DNN can be designed to perform regression.

A trained DNN which provides an optimal configuration of parameters for a given event can be used as an explicit step in the data processing. However, it can also be used only as a research tool, in order to obtain general insights into the correlations between the input and output parameters. In this case, if the specific DNN output values are not used in further analysis, it is not required that the input features are thoroughly controlled for the agreement between data and MC.

B.1.3 Architecture

In most of the DNN applications featured in this thesis, a relatively simple fully connected architecture is used. An example of a fully connected network is illustrated in Figure B.1. In this design, all nodes in a given hidden layer are connected to all nodes in the previous and the next layers. The nodes within the same layer are not connected, and neither are the nodes in non-consecutive layers. The activation function applied to the hidden layers is tanh in all considered applications, as no difference was observed when using other functions.

All networks used in the studies contain either two or three hidden layers with an “inverted pyramid” layout, in which the number of nodes in a given hidden layer is less than or equal than that in the preceding hidden layer. The DNN architectures are optimized by

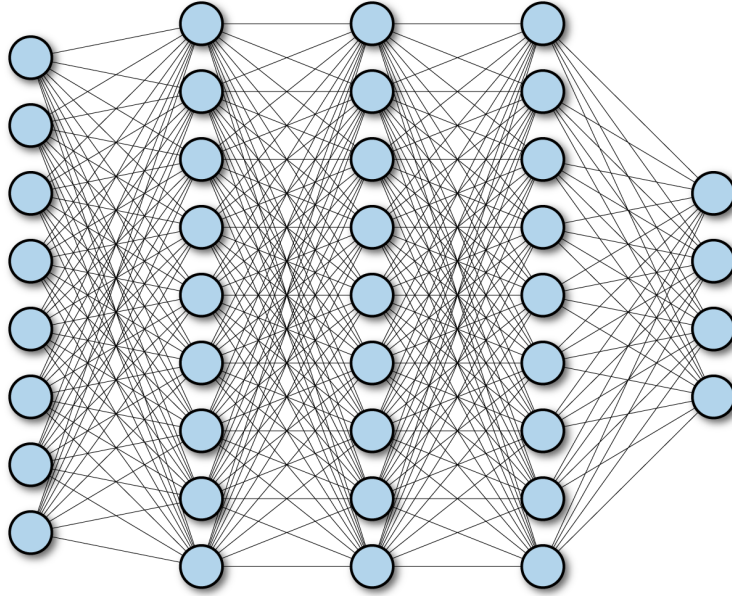


Figure B.1. An example of a fully-connected DNN architecture [168].

performing a grid search, allowing the number of nodes in each hidden layer to be equal to a power of 2 in a reasonable range (32–2048). The optimization is performed for the applications in which the quality of the DNN performance is of a particular importance (Sections 4.6.2, 3.4), but not in the studies where the DNN is used for exploratory purposes (Sections 5.1.2, 5.1.3). In the latter, a generic 3-layer architecture with 64, 32, and 16 nodes in the hidden layers is used, as it is sufficient to demonstrate an improvement over more traditional methods. The implementation of the DNN models was done using KERAS [169] and PYTORCH [143] packages.

B.1.4 Loss functions

The DNN applications in this thesis used a supervised learning approach, in which for every training event the true value of the target parameter is known. Multiple options were considered for the loss function, as listed below. In all cases, the loss function is a function of the DNN prediction y^{pred} and corresponding true values y^{true} , computed for a given batch of N events.

- **Binary cross-entropy (BCE):**

$$L(y_i^{\text{pred}}, y_i^{\text{true}}) = - \sum_{i=1}^N \left(y_i^{\text{true}} \log y_i^{\text{pred}} + (1 - y_i^{\text{true}}) \log(1 - y_i^{\text{pred}}) \right) \quad (\text{B.1})$$

- **Mean squared error (MSE):**

$$L(y_i^{\text{pred}}, y_i^{\text{true}}) = \frac{1}{N} \sum_{i=1}^N (y_i^{\text{pred}} - y_i^{\text{true}})^2 \quad (\text{B.2})$$

- **Significance loss [144]:**

$$\begin{aligned} L(y_i^{\text{pred}}, y_i^{\text{true}}, W_s, W_b) &= (s + b)/b^2, \\ s &= W_s \sum_{i=1}^N y_i^{\text{pred}} \times y_i^{\text{true}}, \\ b &= W_b \sum_{i=1}^N y_i^{\text{pred}} \times (1 - y_i^{\text{true}}), \end{aligned} \quad (\text{B.3})$$

where W_s and W_b are signal and background weights that scale the yields to their expected values. This loss was designed to directly optimize the signal significance.

- **Likelihood loss [145]:**

$$L(y_i^{\text{pred}}, y_i^{\text{true}}, W_b) = \sum_{i=1}^N \left[W_b (1 - y_i^{\text{true}}) (e^{y_i^{\text{pred}}} - 1) - y_i^{\text{pred}} * y_i^{\text{true}} \right] \quad (\text{B.4})$$

This loss function is supposed to explicitly produce the logarithm of the likelihood ratio as a DNN score.

The BCE and MSE losses are the general purpose functions widely used in machine learning applications, while the “significance loss” and “likelihood loss” were specifically designed to optimize the figures of merit specific to HEP analyses. In a study aimed at the signal significance optimization in the VBF channel of the $H \rightarrow \mu\mu$ search (Section 5.2.3), the four loss functions were compared, but no improvement was found when “significance loss” or “likelihood loss” were used. Therefore, in all DNN applications mentioned in this thesis, the cross-entropy loss was chosen.

B.1.5 Cross-validation

An important requirement in the training of machine learning methods is not to use the same data for training, hyperparameter optimization, and application of the trained model, in order to avoid biasing the results. In the DNN applications described in this thesis, this requirement was taken into account by implementing the *k-fold* approach, illustrated in Figure B.2.

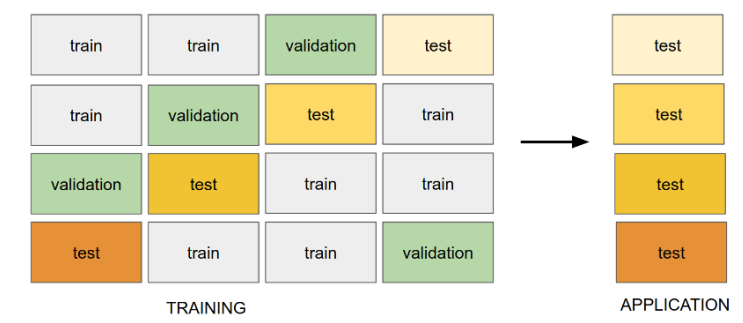


Figure B.2. Splitting into training, validation and evaluation dataset in the *k*-fold procedure with four folds [2].

The training data was split into four subset of equal size, based on the *event number* — a unique identifier of an event within the same data or MC dataset. In each DNN implementation, four models were trained with identical settings, but using different *training folds*, each containing 50% of the training dataset (grey blocks in Figure B.2). The loss and/or other figures of merit used for hyperparameter tuning were evaluated using the *validation folds*, each containing 25% of the dataset (green blocks in Figure B.2), and the final evaluation

was performed using *testing folds*, each containing 25% of the dataset (orange blocks in Figure B.2). Since the event number is fixed for each event, training, tuning, and evaluation of all DNN models were always performed on different events.

B.1.6 Criterion for the best model choice

Although in the MC template-based strategy of the $H \rightarrow \mu\mu$ search the DNN is designed to maximize the separation between signal and background, it has been found that the model corresponding to the global minimum of the loss function does not necessarily produce the most sensitive output score. This may happen when the training datasets do not have exactly the same feature distributions as the datasets used for evaluation. For instance, in our implementation of the VBF $H \rightarrow \mu\mu$ analysis, a mix of signal MC datasets generated with PYTHIA [115] and HERWIG [114] generators is used for training to increase the number of training events, while only the PYTHIA dataset is used in evaluation.

The selection of the best model is done by explicitly evaluating a figure of merit that approximates the model sensitivity. After each training iteration, the model is evaluated on the validation fold of the dataset (25% of events), and the output score distribution is used to evaluate the expected significance. The significance is defined as

$$\sigma = \sqrt{\sum_{\text{bins}} \left(\frac{S_i}{\sqrt{B_i}} \right)^2}, \quad (\text{B.5})$$

where S_i is the expected signal yield in i -th bin, B_i is the expected background yield, and both signal and background yields are computed using events from exactly the same MC samples which are later used in the actual statistical analysis in the template-based category. Equation B.5 is an approximation of *Asimov significance*

$$\sigma_{\text{Asim.}} = \sqrt{\sum_{\text{bins}} \left[(S_i + B_i) \log \left(1 - \frac{S_i}{B_i} \right) - S_i \right]}$$

for $S \ll B$, and it has been checked that this approximation does not have a significant impact on the model selection method.

The binning scheme used in the significance calculation is based on the VBF $H \rightarrow \mu\mu$ signal MC distribution, and configured to not significantly deviate from the binning scheme used in the Run 2 analysis. The number of VBF $H \rightarrow \mu\mu$ events is required to be uniformly distributed in 13 bins ranging from minimal to maximal DNN score values in the validation dataset, and the highest bin boundary is later increased to a very high number to prevent losing the possible overflow events. The event yields in this case are computed using unweighted events, because evaluating simple quantiles is much faster from a computational point of view, as compared to weighted quantiles. This approximation may be one of the reasons for the fluctuations of the significance (see Figure B.3). However, it allows to update the binning at each iteration, which helps to avoid a bias that is introduced when a fixed binning is used.

The evolution of both validation loss and validation significance for one of the training folds is shown in Figure B.3 (left). The training is intentionally performed for an excessive number of iterations (training batches), in order to illustrate the behavior of the figures of merit. It is evident that the values of loss and significance evaluated on the validation dataset are not fully correlated, and the minimum of the loss does not correspond to the maximal significance.

Another illustration of this fact can be seen in Figure B.3 (right), where three trainings with identical settings but different random seeds are shown. While the evolution of the validation loss is very similar for all considered options, the evolution of the expected significance shows considerable differences in behavior, not entirely explained by statistical fluctuations.

The approach of choosing the model with the highest significance has been further validated by training the DNN for 10000 iterations, and evaluating the total analysis sensitivity with a model taken at the point of lowest validation loss, and with a model taken at the point of the highest validation significance. The expected significance for the model with the lowest loss was evaluated to be 1.74σ , while for the model with highest significance it was 1.83σ (5% higher). Since the DNN score templates are produced using the “test” fold of the dataset, which does not overlap with the training and validation folds by design (see Sec-

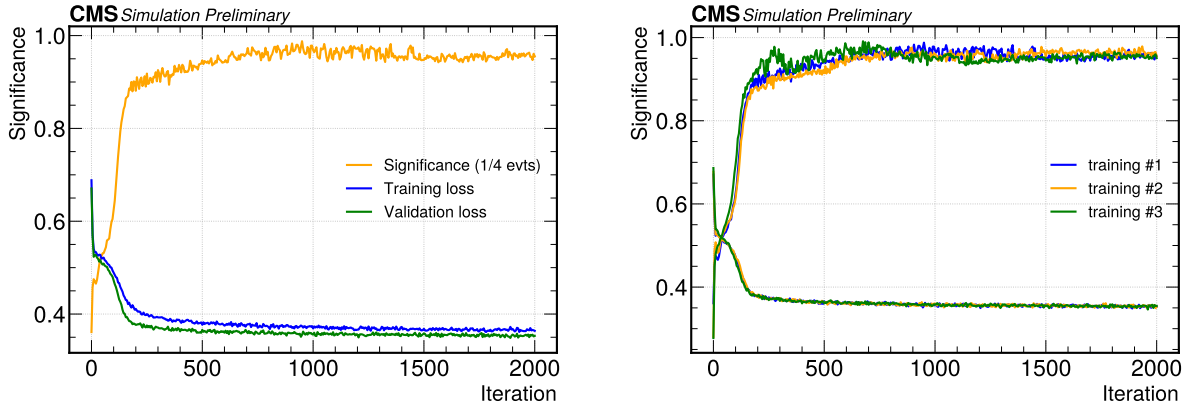


Figure B.3. Left: evolution of the training and validation losses, as well as the expected significance computed for the validation dataset (25% events). Right: validation loss and significance for three trainings with identical settings.

tion B.1.5), this result confirms that the differences in the evolution of loss and significance are not caused by overfitting or statistical fluctuations.

B.2 Boosted Decision Trees (BDT)

Boosted decision tree (BDT) is another machine learning method widely used in high energy physics. Although in general the popularity of BDTs has declined in favor of neural networks, some of the qualities of the BDTs make them still useful for application in HEP analyses.

The BDTs in particle physics are used for classification tasks, distinguishing between two or more classes of events. Instead of computing an analytic function of the input features, as it is done in case of a neural network, a decision tree provides a boundary which splits the dataset into the regions enriched with events of different classes. The boundary is defined using a tree-like approach, illustrated in Figure B.4.

The boundary is updated iteratively, at each step increasing the weights of misidentified events (this procedure is called *boosting*), and the values in the BDT “leaves” are set based on the composition of the events in them (i.e. a sequence of decisions which yields a higher signal-to-background ratio would correspond to a higher output value). In order to decrease

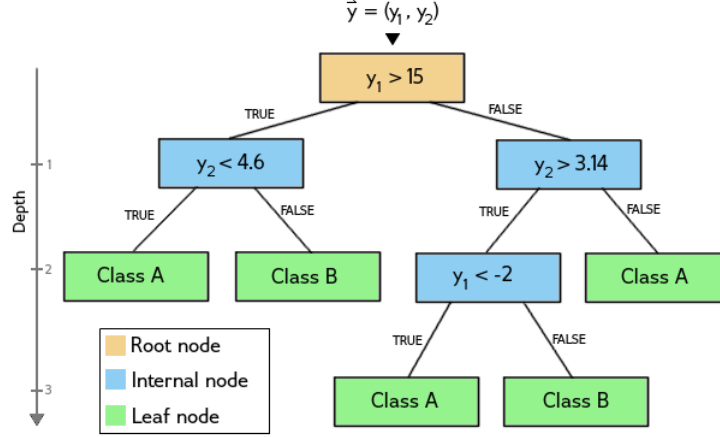


Figure B.4. Graphical representation of an example decision tree [170].

fluctuations in the BDT output, the final score can be obtained by averaging the results yielded by multiple BDT models.

The disadvantage of BDT in comparison to DNN is its limited flexibility due to inability of the BDT to model non-linear correlations with a high level of precision. Consequently, BDT does not perform well in the tasks where the input features are low-level variables such as kinematic properties of individual particles, which do not provide a high discriminating power by themselves, but only in a complex non-linear combination. Deep neural networks are usually show a superior performance in such problems.

However, a benefit of using BDT in the $H \rightarrow \mu\mu$ analysis is related to a more convenient way of weighting the events, as the weighting is already a natural part of the boosting procedure. This quality allows to easily design a BDT model which produces an output score correlated with the dimuon mass resolution in the data-driven categories on the $H \rightarrow \mu\mu$ search (see Section 4.7), which is an integral step in the analysis strategy.

BIBLIOGRAPHY

- [1] CMS Collaboration, “Evidence for Higgs boson decay to a pair of muons”, *JHEP* **01** (2021) 148, doi:[10.1007/JHEP01\(2021\)148](https://doi.org/10.1007/JHEP01(2021)148), arXiv:[2009.04363](https://arxiv.org/abs/2009.04363).
- [2] D. Acosta et al., “Search for Higgs boson decays into muons in proton-proton collisions at 13 TeV using Run 2 data sets”, CMS Analysis Note CMS-AN-2019/124, 2019.
- [3] D. Kondratyev, “copperhead – Columnar Parallel Pythonic framEwork for Run 3 $H \rightarrow \mu\mu$ Decay search”. <https://github.com/Run3HmmAnalysis/copperhead>, 2020–2022.
- [4] D. Kondratyev, S. Piperov, and N. Neumeister, “Implementation and benchmarking of a VBF $H \rightarrow \mu^+\mu^-$ analysis framework using an interactive, parallel Analysis-Facility environment”, CMS Analysis Note CMS-AN-2020/240, 2020.
- [5] D. Kondratyev, A. Purohit, J.-F. Schulte, and N. Neumeister, “Optimization of the outside-in Level-3 muon reconstruction for CMS Muon High Level Trigger”, CMS Analysis Note CMS-AN-2021/156, 2021.
- [6] CMS Collaboration, “The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger”, Technical Design Report CERN-LHCC-2021-007, CMS-TDR-022, CERN, Geneva, 2021. <https://cds.cern.ch/record/2759072>.
- [7] CMS Collaboration, “Snowmass White Paper Contribution: Physics with the Phase-2 ATLAS and CMS Detectors”, CMS Physics Analysis Summary CMS-PAS-FTR-22-001, CERN, Geneva, 2022. <https://cds.cern.ch/record/2806962>.
- [8] S. Dawson et al., “Report of the Topical Group on Higgs Physics for Snowmass 2021: The Case for Precision Higgs Physics”, in *2022 Snowmass Summer Study*. 2022. arXiv:[2209.07510](https://arxiv.org/abs/2209.07510).
- [9] D. Kondratyev, A. Purohit, and N. Neumeister, “Extrapolated measurement precision for Higgs boson properties in the $H \rightarrow \mu\mu$ decay channel at the HL-LHC for Snowmass 2022”, CMS Analysis Note CMS-AN-2022/006, 2022.
- [10] CMS Collaboration, “Prospects for the precise measurement of the Higgs boson properties in the $H \rightarrow \mu\mu$ decay channel at the HL-LHC”, CMS Physics Analysis Summary CMS-PAS-FTR-21-006, 2022. <https://cds.cern.ch/record/2804002>.
- [11] D. Kondratyev, “Studies to improve the MC template-based analysis strategy of the $H \rightarrow \mu\mu$ search in preparation for Run 3”, CMS Analysis Note CMS-AN-2022/145, 2022.

- [12] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Phys. Rev. Lett.* **30** (1973) 1346–1349, [doi:10.1103/PhysRevLett.30.1346](https://doi.org/10.1103/PhysRevLett.30.1346).
- [13] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl. Phys.* **22** (1961) 579–588, [doi:10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [14] 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).
- [15] 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).
- [16] 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).
- [17] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508–509, [doi:10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [18] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321–323, [doi:10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [19] L. Boyle, “Standard model of particle physics, most complete diagram.”, 2014. https://commons.m.wikimedia.org/wiki/File:Standard_Model_Of_Particle_Physics--Most_Complete_Diagram.png.
- [20] Super-Kamiokande Collaboration, “Evidence for oscillation of atmospheric neutrinos”, *Phys. Rev. Lett.* **81** (1998) 1562–1567, [doi:10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562), [arXiv:hep-ex/9807003](https://arxiv.org/abs/hep-ex/9807003).
- [21] SNO Collaboration, “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory”, *Phys. Rev. Lett.* **89** (2002) 011301, [doi:10.1103/PhysRevLett.89.011301](https://doi.org/10.1103/PhysRevLett.89.011301), [arXiv:nuc1-ex/0204008](https://arxiv.org/abs/nuc1-ex/0204008).
- [22] KATRIN Collaboration, “Direct neutrino-mass measurement with sub-electronvolt sensitivity”, *Nature Phys.* **18** (2022), no. 2, 160–166, [doi:10.1038/s41567-021-01463-1](https://doi.org/10.1038/s41567-021-01463-1), [arXiv:2105.08533](https://arxiv.org/abs/2105.08533).
- [23] Particle Data Group, “Review of Particle Physics”, *Phys. Rev. D* **98** (2018), no. 3, 030001, [doi:10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001).
- [24] 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).
- [25] T. Nakano and K. Nishijima, “Charge Independence for V-particles”, *Prog. Theor. Phys.* **10** (1953) 581–582, [doi:10.1143/PTP.10.581](https://doi.org/10.1143/PTP.10.581).

- [26] M. Gell-Mann, “The interpretation of the new particles as displaced charge multiplets”, *Nuovo Cim.* **4** (1956), no. S2, 848–866, [doi:10.1007/BF02748000](#).
- [27] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* **13** (1964) 585–587, [doi:10.1103/PhysRevLett.13.585](#).
- [28] P. W. Higgs, “Spontaneous Symmetry Breakdown without Massless Bosons”, *Phys. Rev.* **145** (1966) 1156–1163, [doi:10.1103/PhysRev.145.1156](#).
- [29] T. W. B. Kibble, “Symmetry Breaking in Non-Abelian Gauge Theories”, *Phys. Rev.* **155** (1967) 1554–1561, [doi:10.1103/PhysRev.155.1554](#).
- [30] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, “A Historical Profile of the Higgs Boson”, pp. 255–274. World Scientific, 2016. [arXiv:1504.07217](#). [doi:10.1142/9789814733519_0014](#).
- [31] J. Goldstone, “Field Theories with Superconductor Solutions”, *Nuovo Cim.* **19** (1961) 154–164, [doi:10.1007/BF02812722](#).
- [32] G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees, “Formation of Galaxies and Large Scale Structure with Cold Dark Matter”, *Nature* **311** (1984) 517–525, [doi:10.1038/311517a0](#).
- [33] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, “The Evolution of Large Scale Structure in a Universe Dominated by Cold Dark Matter”, *Astrophys. J.* **292** (1985) 371–394, [doi:10.1086/163168](#).
- [34] D. Clowe et al., “A Direct Empirical Proof of the Existence of Dark Matter”, *Astrophys. J. Lett.* **648** (2006) 109–113, [doi:10.1086/508162](#), [arXiv:astro-ph/0608407](#).
- [35] Supernova Cosmology Project, “Measurements of Ω and Λ from 42 High-Redshift Supernovae”, *Astrophys. J.* **517** (1999) 565–586, [doi:10.1086/307221](#), [arXiv:astro-ph/9812133](#).
- [36] S. S. Team, “Observational evidence from supernovae for an accelerating universe and a cosmological constant”, *Astron. J.* **116** (1998) 1009–1038, [doi:10.1086/300499](#), [arXiv:astro-ph/9805201](#).
- [37] A. G. Cohen, A. De Rujula, and S. L. Glashow, “A Matter – antimatter universe?”, *Astrophys. J.* **495** (1998) 539–549, [doi:10.1086/305328](#), [arXiv:astro-ph/9707087](#).
- [38] S. Weinberg, “The U(1) Problem”, *Phys. Rev. D* **11** (1975) 3583–3593, [doi:10.1103/PhysRevD.11.3583](#).

- [39] N. Cabibbo, “Unitary Symmetry and Leptonic Decays”, *Phys. Rev. Lett.* **10** (1963) 531–533, [doi:10.1103/PhysRevLett.10.531](https://doi.org/10.1103/PhysRevLett.10.531).
- [40] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, *Prog. Theor. Phys.* **49** (1973) 652–657, [doi:10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [41] H. E. Haber and G. L. Kane, “The Search for Supersymmetry: Probing Physics Beyond the Standard Model”, *Phys. Rept.* **117** (1985) 75–263, [doi:10.1016/0370-1573\(85\)90051-1](https://doi.org/10.1016/0370-1573(85)90051-1).
- [42] “LHC Higgs Cross Section Working group”.
<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>. Accessed: 2022-09-05.
- [43] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”, [doi:10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002), [arXiv:1610.07922](https://arxiv.org/abs/1610.07922).
- [44] C. Anastasiou et al., “High precision determination of the gluon fusion Higgs boson cross-section at the LHC”, *JHEP* **05** (2016) 058, [doi:10.1007/JHEP05\(2016\)058](https://doi.org/10.1007/JHEP05(2016)058), [arXiv:1602.00695](https://arxiv.org/abs/1602.00695).
- [45] CMS Collaboration, “Observation of $t\bar{t}H$ production”, *Phys. Rev. Lett.* **120** (2018), no. 23, 231801, [doi:10.1103/PhysRevLett.120.231801](https://doi.org/10.1103/PhysRevLett.120.231801), [arXiv:1804.02610](https://arxiv.org/abs/1804.02610).
- [46] ATLAS Collaboration, “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector”, *Phys. Lett. B* **784** (2018) 173–191, [doi:10.1016/j.physletb.2018.07.035](https://doi.org/10.1016/j.physletb.2018.07.035), [arXiv:1806.00425](https://arxiv.org/abs/1806.00425).
- [47] CMS Collaboration, “Observation of Higgs boson decay to bottom quarks”, *Phys. Rev. Lett.* **121** (2018), no. 12, 121801, [doi:10.1103/PhysRevLett.121.121801](https://doi.org/10.1103/PhysRevLett.121.121801), [arXiv:1808.08242](https://arxiv.org/abs/1808.08242).
- [48] ATLAS Collaboration, “Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector”, *Phys. Lett. B* **786** (2018) 59–86, [doi:10.1016/j.physletb.2018.09.013](https://doi.org/10.1016/j.physletb.2018.09.013), [arXiv:1808.08238](https://arxiv.org/abs/1808.08238).
- [49] CMS Collaboration, “Measurement of the inclusive and differential Higgs boson production cross sections in the leptonic WW decay mode at $\sqrt{s} = 13$ TeV”, *JHEP* **03** (2021) 003, [doi:10.1007/JHEP03\(2021\)003](https://doi.org/10.1007/JHEP03(2021)003), [arXiv:2007.01984](https://arxiv.org/abs/2007.01984).
- [50] CMS Collaboration, “Measurement of the inclusive and differential Higgs boson production cross sections in the decay mode to a pair of τ leptons in pp collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. Lett.* **128** (2022), no. 8, 081805, [doi:10.1103/PhysRevLett.128.081805](https://doi.org/10.1103/PhysRevLett.128.081805), [arXiv:2107.11486](https://arxiv.org/abs/2107.11486).

- [51] ATLAS Collaboration, “Measurements of Higgs boson production cross-sections in the $H \rightarrow \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *JHEP* **08** (2022) 175, [doi:10.1007/JHEP08\(2022\)175](https://doi.org/10.1007/JHEP08(2022)175), [arXiv:2201.08269](https://arxiv.org/abs/2201.08269).
- [52] CMS Collaboration, “A search for the standard model Higgs boson decaying to charm quarks”, *JHEP* **03** (2020) 131, [doi:10.1007/JHEP03\(2020\)131](https://doi.org/10.1007/JHEP03(2020)131), [arXiv:1912.01662](https://arxiv.org/abs/1912.01662).
- [53] ATLAS Collaboration, “Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector”, *Eur. Phys. J. C* **82** (2022) 717, [doi:10.1140/epjc/s10052-022-10588-3](https://doi.org/10.1140/epjc/s10052-022-10588-3), [arXiv:2201.11428](https://arxiv.org/abs/2201.11428).
- [54] CMS Collaboration, “Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **81** (2021), no. 6, 488, [doi:10.1140/epjc/s10052-021-09200-x](https://doi.org/10.1140/epjc/s10052-021-09200-x), [arXiv:2103.04956](https://arxiv.org/abs/2103.04956).
- [55] ATLAS Collaboration, “Measurements of the Higgs boson inclusive and differential fiducial cross sections in the 4ℓ decay channel at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **80** (2020), no. 10, 942, [doi:10.1140/epjc/s10052-020-8223-0](https://doi.org/10.1140/epjc/s10052-020-8223-0), [arXiv:2004.03969](https://arxiv.org/abs/2004.03969).
- [56] CMS Collaboration, “Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2021) 027, [doi:10.1007/JHEP07\(2021\)027](https://doi.org/10.1007/JHEP07(2021)027), [arXiv:2103.06956](https://arxiv.org/abs/2103.06956).
- [57] ATLAS Collaboration, “Measurements of the Higgs boson inclusive and differential fiducial cross-sections in the diphoton decay channel with pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *JHEP* **08** (2022) 027, [doi:10.1007/JHEP08\(2022\)027](https://doi.org/10.1007/JHEP08(2022)027), [arXiv:2202.00487](https://arxiv.org/abs/2202.00487).
- [58] CMS Collaboration, “Search for Higgs boson decays to a Z boson and a photon in proton-proton collisions at $\sqrt{s} = 13$ TeV”, [arXiv:2204.12945](https://arxiv.org/abs/2204.12945).
- [59] ATLAS Collaboration, “A search for the $Z\gamma$ decay mode of the Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Lett. B* **809** (2020) 135754, [doi:10.1016/j.physletb.2020.135754](https://doi.org/10.1016/j.physletb.2020.135754), [arXiv:2005.05382](https://arxiv.org/abs/2005.05382).
- [60] ATLAS Collaboration, “A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector”, *Phys. Lett. B* **812** (2021) 135980, [doi:10.1016/j.physletb.2020.135980](https://doi.org/10.1016/j.physletb.2020.135980), [arXiv:2007.07830](https://arxiv.org/abs/2007.07830).
- [61] 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).

- [62] 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).
- [63] CMS Collaboration, “Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector”, *Phys. Lett. B* **779** (2018) 283–316, [doi:10.1016/j.physletb.2018.02.004](https://doi.org/10.1016/j.physletb.2018.02.004), [arXiv:1708.00373](https://arxiv.org/abs/1708.00373).
- [64] ATLAS Collaboration, “Cross-section measurements of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Rev. D* **99** (2019) 072001, [doi:10.1103/PhysRevD.99.072001](https://doi.org/10.1103/PhysRevD.99.072001), [arXiv:1811.08856](https://arxiv.org/abs/1811.08856).
- [65] CMS Collaboration, “Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production”, *Nat. Phys.* **18** (2022) 1329–1334, [doi:10.1038/s41567-022-01682-0](https://doi.org/10.1038/s41567-022-01682-0), [arXiv:2202.06923](https://arxiv.org/abs/2202.06923).
- [66] UA1 Collaboration, “Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c² at the CERN SPS Collider”, *Phys. Lett. B* **126** (1983) 398–410, [doi:10.1016/0370-2693\(83\)90188-0](https://doi.org/10.1016/0370-2693(83)90188-0).
- [67] UA2 Collaboration, “Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN $\bar{p}p$ Collider”, *Phys. Lett. B* **129** (1983) 130–140, [doi:10.1016/0370-2693\(83\)90744-X](https://doi.org/10.1016/0370-2693(83)90744-X).
- [68] UA1 Collaboration, “Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV”, *Phys. Lett. B* **122** (1983) 103–116, [doi:10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2).
- [69] UA2 Collaboration, “Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider”, *Phys. Lett. B* **122** (1983) 476–485, [doi:10.1016/0370-2693\(83\)91605-2](https://doi.org/10.1016/0370-2693(83)91605-2).
- [70] O. S. Brüning et al., “LHC Design Report”. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. [doi:10.5170/CERN-2004-003-V-1](https://doi.org/10.5170/CERN-2004-003-V-1).
- [71] “LHC Machine”, *JINST* **3** (2008) S08001, [doi:10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [72] “LEP design report”, technical report, Geneva, 1984. <https://cds.cern.ch/record/102083>.
- [73] 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).
- [74] 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).

- [75] 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).
- [76] 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).
- [77] P. Mouche, “Overall view of the LHC. Vue d’ensemble du LHC”, 2014. <https://cds.cern.ch/record/1708847>.
- [78] D. J. Warner, “Project study for a new 50 MeV linear accelerator for the C. P. S”, Technical Report CERN-MPS-LINP-73-1, CERN, Geneva, 1973. <https://cds.cern.ch/record/414071>.
- [79] L. Arnaudon et al., “Linac4 technical design report”, Technical Report CERN-AB-2006-084, 2006. <https://cds.cern.ch/record/1004186>.
- [80] F. Marcastel, “CERN’s Accelerator Complex. La chaîne des accélérateurs du CERN”, 2013. <https://cds.cern.ch/record/1621583>.
- [81] “CMS Luminosity - Public Results”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>. Accessed: 2022-07-31.
- [82] CMS Collaboration, “Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2018) 161, [doi:10.1007/JHEP07\(2018\)161](https://doi.org/10.1007/JHEP07(2018)161), [arXiv:1802.02613](https://arxiv.org/abs/1802.02613).
- [83] “Longer Term LHC Schedule”. <https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm>. Accessed: 2022-07-31.
- [84] CMS Collaboration, “Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays”, *JINST* **5** (2010) T03021, [doi:10.1088/1748-0221/5/03/T03021](https://doi.org/10.1088/1748-0221/5/03/T03021), [arXiv:0910.5530](https://arxiv.org/abs/0910.5530).
- [85] CMS Collaboration, “The CMS tracker system project: Technical Design Report”, Technical Design Report CERN-LHCC-98-006, CMS-TDR-5, Geneva, 1997. <https://cds.cern.ch/record/368412>.
- [86] CMS Collaboration, “The CMS tracker: addendum to the Technical Design Report”, Technical Design Report CERN-LHCC-2000-016, CMS-TDR-5-add-1, Geneva, 2000. <https://cds.cern.ch/record/490194>.
- [87] CMS Collaboration, “Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays”, *JINST* **5** (2010) T03009, [doi:10.1088/1748-0221/5/03/T03009](https://doi.org/10.1088/1748-0221/5/03/T03009), [arXiv:0910.2505](https://arxiv.org/abs/0910.2505).

- [88] CMS Tracker Group, “The CMS Phase-1 Pixel Detector Upgrade”, *JINST* **16** (2021), no. 02, P02027, [doi:10.1088/1748-0221/16/02/P02027](https://doi.org/10.1088/1748-0221/16/02/P02027), [arXiv:2012.14304](https://arxiv.org/abs/2012.14304).
- [89] M. Dragicevic et al., “Test Beam Performance Measurements for the Phase I Upgrade of the CMS Pixel Detector”, *JINST* **12** (2017), no. 05, P05022, [doi:10.1088/1748-0221/12/05/P05022](https://doi.org/10.1088/1748-0221/12/05/P05022), [arXiv:1706.00222](https://arxiv.org/abs/1706.00222).
- [90] CMS Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report”, Technical Design Report CERN-LHCC-97-033, CMS-TDR-4, Geneva, 1997. <https://cds.cern.ch/record/349375>.
- [91] CMS Collaboration, “The CMS ECAL performance with examples”, *JINST* **9** (2014) C02008, [doi:10.1088/1748-0221/9/02/C02008](https://doi.org/10.1088/1748-0221/9/02/C02008).
- [92] CMS Collaboration, “The CMS hadron calorimeter project: Technical Design Report”, Technical Design Report CERN-LHCC-97-031, CMS-TDR-2, Geneva, 1997. <https://cds.cern.ch/record/357153>.
- [93] CMS Collaboration, “Performance of CMS Hadron Calorimeter Timing and Synchronization using Test Beam, Cosmic Ray, and LHC Beam Data”, *JINST* **5** (2010) T03013, [doi:10.1088/1748-0221/5/03/T03013](https://doi.org/10.1088/1748-0221/5/03/T03013), [arXiv:0911.4877](https://arxiv.org/abs/0911.4877).
- [94] USCMS, ECAL/HCAL Collaboration, “The CMS barrel calorimeter response to particle beams from 2 GeV/c to 350 GeV/c”, *Eur. Phys. J. C* **60** (2009) 359–373, [doi:10.1140/epjc/s10052-009-0959-5](https://doi.org/10.1140/epjc/s10052-009-0959-5).
- [95] CMS Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter”, Technical Design Report CERN-LHCC-2017-023, CMS-TDR-019, CERN, Geneva, 2017. [doi:10.17181/CERN.IV8M.1JY2](https://doi.org/10.17181/CERN.IV8M.1JY2).
- [96] CMS Collaboration, “The CMS muon project: Technical Design Report”, Technical Design Report CERN-LHCC-97-032, CMS-TDR-3, Geneva, 1997. <https://cds.cern.ch/record/343814>.
- [97] 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).
- [98] “Summaries of CMS cross section measurements”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined>. Accessed: 2022-08-12.
- [99] CMS Collaboration, “CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems”, Technical Design Report CERN-LHCC-2000-038, 2000. <https://cds.cern.ch/record/706847>.

- [100] CMS Trigger, Data Acquisition Group Collaboration, “The CMS high level trigger”, *Eur. Phys. J. C* **46** (2006) 605–667, [doi:10.1140/epjc/s2006-02495-8](#), [arXiv:hep-ex/0512077](#).
- [101] 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](#).
- [102] 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](#), [arXiv:1405.6569](#).
- [103] A. Strandlie and W. Wittek, “Derivation of jacobians for the propagation of covariance matrices of track parameters in homogeneous magnetic fields”, *Nuclear Instruments and Methods in Physics Research Section A* **566** (2006) 687–698, [doi:10.1016/j.nima.2006.07.032](#).
- [104] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting”, *Nuclear Instruments and Methods in Physics Research Section A* **262** (1987) 444–450, [doi:10.1016/0168-9002\(87\)90887-4](#).
- [105] S. Chatrchyan et al., “Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ tev with the CMS experiment”, *JHEP* **2011** (2011), no. 10, [doi:10.1007/jhep10\(2011\)132](#).
- [106] CMS Collaboration, “Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **16** (2021) P07001, [doi:10.1088/1748-0221/16/07/P07001](#), [arXiv:2102.04790](#).
- [107] K. Pearson, “On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling”, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **50** (1900), no. 302, 157–175, [doi:10.1080/14786440009463897](#).
- [108] S. Folgueras, N. Neumeister, and B. Mahakud, “HLT muon reconstruction using iterative tracking”, CMS Analysis Note CMS-AN-2017/185, 2019.
- [109] J. Han and C. Moraga, “The influence of the sigmoid function parameters on the speed of backpropagation learning”, in *From Natural to Artificial Neural Computation*, J. Mira and F. Sandoval, eds., pp. 195–201. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995.
- [110] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization”, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA*,

- USA, May 7-9, 2015, *Conference Track Proceedings*, Y. Bengio and Y. LeCun, eds. 2015. <http://arxiv.org/abs/1412.6980>.
- [111] T. O'Malley et al., “Keras Tuner”. <https://github.com/keras-team/keras-tuner>, 2019.
 - [112] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **06** (2011) 128, [doi:10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128), [arXiv:1106.0522](https://arxiv.org/abs/1106.0522).
 - [113] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, [doi:10.1088/1126-6708/2007/11/070](https://doi.org/10.1088/1126-6708/2007/11/070), [arXiv:0709.2092](https://arxiv.org/abs/0709.2092).
 - [114] T. Sjostrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1”, *Comput. Phys. Commun.* **178** (2008) 852–867, [doi:10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036), [arXiv:0710.3820](https://arxiv.org/abs/0710.3820).
 - [115] J. Bellm et al., “Herwig 7.0/Herwig++ 3.0 release note”, *Eur. Phys. J. C* **76** (2016), no. 4, 196, [doi:10.1140/epjc/s10052-016-4018-8](https://doi.org/10.1140/epjc/s10052-016-4018-8), [arXiv:1512.01178](https://arxiv.org/abs/1512.01178).
 - [116] GEANT4 Collaboration, “GEANT4: A Simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, [doi:10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
 - [117] J. Allison et al., “Geant4 developments and applications”, *IEEE Trans. Nucl. Sci.* **53** (2006) 270, [doi:10.1109/TNS.2006.869826](https://doi.org/10.1109/TNS.2006.869826).
 - [118] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018), no. 05, P05011, [doi:10.1088/1748-0221/13/05/P05011](https://doi.org/10.1088/1748-0221/13/05/P05011), [arXiv:1712.07158](https://arxiv.org/abs/1712.07158).
 - [119] A. Bodek et al., “Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments”, *Eur. Phys. J. C* **72** (2012) 2194, [doi:10.1140/epjc/s10052-012-2194-8](https://doi.org/10.1140/epjc/s10052-012-2194-8), [arXiv:1208.3710](https://arxiv.org/abs/1208.3710).
 - [120] Particle Data Group, “Review of Particle Physics”, *Progress of Theoretical and Experimental Physics* **2020** (08, 2020) [doi:10.1093/ptep/ptaa104](https://doi.org/10.1093/ptep/ptaa104).
 - [121] X. Zuo, “A Search for the Standard Model Higgs Boson Decaying into Two Muons in the Vector Boson Associated Production Mode at the CMS Experiment”. PhD thesis, University of Florida, 2021. <https://cds.cern.ch/record/2790970>.
 - [122] CMS Collaboration, “CMS luminosity measurements for the 2016 data taking period”, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.
 - [123] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

- [124] CMS Collaboration, “CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV”, CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.
- [125] R. J. Barlow and C. Beeston, “Fitting using finite Monte Carlo samples”, *Comput. Phys. Commun.* **77** (1993) 219–228, [doi:10.1016/0010-4655\(93\)90005-W](https://doi.org/10.1016/0010-4655(93)90005-W).
- [126] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies, “Handling uncertainties in background shapes: the discrete profiling method”, *JINST* **10** (2015), no. 04, P04015, [doi:10.1088/1748-0221/10/04/P04015](https://doi.org/10.1088/1748-0221/10/04/P04015), [arXiv:1408.6865](https://arxiv.org/abs/1408.6865).
- [127] D. L. Rainwater, R. Szalapski, and D. Zeppenfeld, “Probing color singlet exchange in $Z + \text{two jet}$ events at the CERN LHC”, *Phys. Rev. D* **54** (1996) 6680–6689, [doi:10.1103/PhysRevD.54.6680](https://doi.org/10.1103/PhysRevD.54.6680), [arXiv:hep-ph/9605444](https://arxiv.org/abs/hep-ph/9605444).
- [128] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2017.
- [129] J. C. Collins and D. E. Soper, “Angular Distribution of Dileptons in High-Energy Hadron Collisions”, *Phys. Rev. D* **16** (1977) 2219, [doi:10.1103/PhysRevD.16.2219](https://doi.org/10.1103/PhysRevD.16.2219).
- [130] 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).
- [131] H. Voss, A. Hocker, J. Stelzer, and F. Tegenfeldt, “TMVA, the Toolkit for Multivariate Data Analysis with ROOT”, *PoS ACAT* (2007) 040, [doi:10.22323/1.050.0040](https://doi.org/10.22323/1.050.0040).
- [132] CMS Collaboration, “Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. Lett.* **122** (2019), no. 2, 021801, [doi:10.1103/PhysRevLett.122.021801](https://doi.org/10.1103/PhysRevLett.122.021801), [arXiv:1807.06325](https://arxiv.org/abs/1807.06325).
- [133] M. J. Oreglia, “A study of the reactions $\psi' \rightarrow \gamma\gamma\psi$ ”. PhD thesis, Stanford University, 1980. SLAC Report SLAC-R-236, <http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-236.pdf>.
- [134] ATLAS Collaboration, “Search for the dimuon decay of the Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Rev. Lett.* **119** (2017), no. 5, 051802, [doi:10.1103/PhysRevLett.119.051802](https://doi.org/10.1103/PhysRevLett.119.051802), [arXiv:1705.04582](https://arxiv.org/abs/1705.04582).
- [135] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in FEWZ”, *Phys. Rev. D* **86** (2012) 094034, [doi:10.1103/PhysRevD.86.094034](https://doi.org/10.1103/PhysRevD.86.094034), [arXiv:1208.5967](https://arxiv.org/abs/1208.5967).

- [136] CMS Collaboration, “Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s}=13$ TeV”, *Phys. Rev. Lett.* **122** (2019), no. 13, 132003, [doi:10.1103/PhysRevLett.122.132003](https://doi.org/10.1103/PhysRevLett.122.132003), [arXiv:1812.05900](https://arxiv.org/abs/1812.05900).
- [137] ATLAS and CMS Collaborations, and the LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, 2011. <https://cds.cern.ch/record/1379837>.
- [138] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435, [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).
- [139] A. L. Read, “Presentation of search results: the cls technique”, *J. Phys. G* **28** (2002) 2693, [doi:10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [140] CMS Collaboration, “Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC”, *Phys. Lett. B* **744** (2015) 184, [doi:10.1016/j.physletb.2015.03.048](https://doi.org/10.1016/j.physletb.2015.03.048), [arXiv:1410.6679](https://arxiv.org/abs/1410.6679).
- [141] CMS Collaboration, “Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s}=13$ TeV”, *Eur. Phys. J. C* **79** (2019), no. 5, 421, [doi:10.1140/epjc/s10052-019-6909-y](https://doi.org/10.1140/epjc/s10052-019-6909-y), [arXiv:1809.10733](https://arxiv.org/abs/1809.10733).
- [142] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties”, (2013). [arXiv:1307.1347](https://arxiv.org/abs/1307.1347).
- [143] A. Paszke et al., “Pytorch: An imperative style, high-performance deep learning library”, in *Advances in Neural Information Processing Systems 32*, H. Wallach et al., eds., pp. 8024–8035. Curran Associates, Inc., 2019. <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.
- [144] A. Elwood, D. Krücker, and M. Shchedrolosiev, “Direct optimization of the discovery significance in machine learning for new physics searches in particle colliders”, *J. Phys. Conf. Ser.* **1525** (2020) 012110, [doi:10.1088/1742-6596/1525/1/012110](https://doi.org/10.1088/1742-6596/1525/1/012110).
- [145] R. T. D’Agnolo and A. Wulzer, “Learning New Physics from a Machine”, *Phys. Rev. D* **99** (2019), no. 1, 015014, [doi:10.1103/PhysRevD.99.015014](https://doi.org/10.1103/PhysRevD.99.015014), [arXiv:1806.02350](https://arxiv.org/abs/1806.02350).
- [146] D. Contardo et al., “Technical Proposal for the Phase-II Upgrade of the CMS Detector”, Technical Design Report CERN-LHCC-2015-010, LHCC-P-008, CMS-TDR-15-02, Geneva, 2015. <https://cds.cern.ch/record/2020886>.

- [147] F. Gianotti et al., “Physics potential and experimental challenges of the LHC luminosity upgrade”, *Eur. Phys. J. C* **39** (2002) 293–333. 62 p,
[doi:10.1140/epjc/s2004-02061-6](https://doi.org/10.1140/epjc/s2004-02061-6).
- [148] CMS Collaboration, “Projected Performance of an Upgraded CMS Detector at the LHC and HL-LHC: Contribution to the Snowmass Process”, Technical Report CMS-NOTE-13-002, CERN, Geneva, 2013. [arXiv:1307.7135](https://arxiv.org/abs/1307.7135).
- [149] S. Dawson et al., “Higgs Working Group Report of the Snowmass 2013 Community Planning Study”, technical report, 2013. [arXiv:1310.8361](https://arxiv.org/abs/1310.8361).
- [150] CMS Collaboration, “Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC”, *Phys. Lett. B* **744** (2015) 184–207,
[doi:10.1016/j.physletb.2015.03.048](https://doi.org/10.1016/j.physletb.2015.03.048), [arXiv:1410.6679](https://arxiv.org/abs/1410.6679).
- [151] CMS Collaboration, “Search for the standard model Higgs boson in the dimuon decay channel in pp collisions at $\sqrt{s}=7$ and 8 TeV”, CMS Physics Analysis Summary CMS-PAS-HIG-13-007, CERN, Geneva, 2013. <https://cds.cern.ch/record/1606831>.
- [152] J. Hugon, “Search for the Standard Model Higgs Boson Decaying to $\mu^+\mu^-$ in pp Collisions at $\sqrt{s}=7$ and 8 TeV with the CMS Detector”, in *2nd Large Hadron Collider Physics Conference*. 2014. [arXiv:1409.0839](https://arxiv.org/abs/1409.0839).
- [153] A. Dainese et al., “Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC”, Technical Report CERN-2019-007, Geneva, Switzerland, 2019.
[doi:10.23731/CYRM-2019-007](https://doi.org/10.23731/CYRM-2019-007).
- [154] CMS Collaboration, “The Phase-2 Upgrade of the CMS Tracker”, Technical Design Report CERN-LHCC-2017-009, CMS-TDR-014, CERN, Geneva, 2017.
<https://cds.cern.ch/record/2272264>.
- [155] “Recommendations for Systematic Uncertainties at HL-LHC”.
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/YR2018Systematics>. Accessed: 2022-08-02.
- [156] CMS Collaboration, “The Phase-2 Upgrade of the CMS Barrel Calorimeters Technical Design Report”, Technical Design Report CERN-LHCC-2017-011, 2017.
<https://cds.cern.ch/record/2283187>.
- [157] CMS Collaboration, “The Phase-2 Upgrade of the CMS Muon Detectors”, Technical Design Report CERN-LHCC-2017-012, CMS-TDR-016, CERN, Geneva, 2017.
<https://cds.cern.ch/record/2283189>.
- [158] DELPHES 3 Collaboration, “DELPHES 3, A modular framework for fast simulation of a generic collider experiment”, *JHEP* **02** (2014) 057,
[doi:10.1007/JHEP02\(2014\)057](https://doi.org/10.1007/JHEP02(2014)057), [arXiv:1307.6346](https://arxiv.org/abs/1307.6346).

- [159] “Delphes instructions for Snowmass2021”.
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/DelphesInstructions>. Accessed:
 2022-07-31.
- [160] A. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, “How bright is the proton? A precise determination of the photon parton distribution function”, *Phys. Rev. Lett.* **117** (2016) 242002, [doi:10.1103/PhysRevLett.117.242002](https://doi.org/10.1103/PhysRevLett.117.242002), [arXiv:1607.04266](https://arxiv.org/abs/1607.04266).
- [161] B. P. Roe et al., “Boosted decision trees, an alternative to artificial neural networks”, *Nucl. Instrum. Meth. A* **543** (2005), no. 2-3, 577–584,
[doi:10.1016/j.nima.2004.12.018](https://doi.org/10.1016/j.nima.2004.12.018), [arXiv:physics/0408124](https://arxiv.org/abs/physics/0408124).
- [162] Y. Bengio, “Learning Deep Architectures for AI”, *Foundations* **2** (2009) 1–55,
[doi:10.1561/22000000006](https://doi.org/10.1561/22000000006).
- [163] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning”, *Nature* **521** (2015) 436–44,
[doi:10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [164] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning”. MIT Press, 2016.
<http://www.deeplearningbook.org>.
- [165] D. Guest, K. Cranmer, and D. Whiteson, “Deep Learning and its Application to LHC Physics”, *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 161–181,
[doi:10.1146/annurev-nucl-101917-021019](https://doi.org/10.1146/annurev-nucl-101917-021019), [arXiv:1806.11484](https://arxiv.org/abs/1806.11484).
- [166] J. Neyman and E. S. Pearson, “On the Problem of the Most Efficient Tests of Statistical Hypotheses”, *Phil. Trans. Roy. Soc. Lond. A* **231** (1933), no. 694-706, 289–337, [doi:10.1098/rsta.1933.0009](https://doi.org/10.1098/rsta.1933.0009).
- [167] B. Nachman, “A guide for deploying Deep Learning in LHC searches: How to achieve optimality and account for uncertainty”, *SciPost Phys.* **8** (2020) 090,
[doi:10.21468/SciPostPhys.8.6.090](https://doi.org/10.21468/SciPostPhys.8.6.090), [arXiv:1909.03081](https://arxiv.org/abs/1909.03081).
- [168] B. Ramsundar and R. Zadeh, “TensorFlow for Deep Learning: From Linear Regression to Reinforcement Learning”. O’Reilly Media, 2018. ISBN 9781491980453.
- [169] F. Chollet et al., “Keras”, 2015. <https://github.com/fchollet/keras>.
- [170] S. Caron et al., “The BSM-AI project: SUSY-AI—generalizing LHC limits on supersymmetry with machine learning”, *The European Physical Journal C* **77** (2016) [doi:10.1140/epjc/s10052-017-4814-9](https://doi.org/10.1140/epjc/s10052-017-4814-9).