

ABSTRACT

STUDIES OF VECTOR BOSON SCATTERING IN THE SEMILEPTONIC ZV CHANNEL WITH THE CMS DETECTOR AT $\sqrt{s} = 13$ TEV

Ramanpreet Singh, Ph.D.

Department of Physics

Northern Illinois University, 2022

Vishnu Zutshi, Director

This dissertation reports on the search for the vector boson scattering (VBS) in semileptonic ZV channel, where the Z boson decays to a pair of leptons (electrons or muons), and the other vector boson V (either W or Z) decays hadronically to a pair of jets, in association of two large pseudorapidity jets i.e. $ZVjj \rightarrow l^+l^-jjjj$, $l = e$ or μ . The search used 137 fb^{-1} of proton-proton collision data collected by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) with center-of-mass energy (\sqrt{s}) of 13 TeV from 2016 to 2018. VBS is a key process in understanding the nature of electroweak symmetry breaking (EWSB) in the framework of Standard Model (SM).

Studies and instrumentation done towards the future detector upgrade to the CMS experiment which will replace the current endcap Hadronic Calorimeter (HCAL) and Electromagnetic Calorimeter (ECAL) with High Granularity Calorimeter (HGCAL) during Long Shutdown 3 (LS3) are also reported. Optimal configuration of scintillator tiles coupled to silicon photomultipliers (SiPMs) in the HGCAL is measured and suggested using testbeam measurement of scintillator tiles and noise measurement of SiPMs. Prototyping of automated wrapping of scintillator tiles in Enhanced Specular Reflector (ESR) film are also discussed.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

SUMMER 2022

**STUDIES OF VECTOR BOSON SCATTERING IN THE SEMILEPTONIC
ZV CHANNEL WITH THE CMS DETECTOR AT $\sqrt{s} = 13$ TEV**

BY

RAMANPREET SINGH
© 2022 Ramanpreet Singh

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

Dissertation Director:
Vishnu Zutshi

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Vishnu Zutshi for giving me opportunities to make this work possible. Your mentorship and guidance over the last six years has contributed to my growth as a scientist and as a person. Thank you for all the discussions and being patient with me. I would also like to thank my committee members Michael (Mike) Eads, Gerald (Jerry) Blazey, and Jeffrey Berryhill for providing feedback on my dissertation.

For the analysis, I would like to thank Michael (Mike) Eads, for providing guidance and suggestions throughout the analysis. I would also like thank Jeffrey Berryhill, Aram Apyan, Jay Lawhorn and CMS collaboration for their help and guidance.

I would like to thank Alexandre (Sasha) Dychkant for teaching me various lab equipments and sharing his experimental knowledge. I would also like to thank Iman Salehinia, Nicholas Pohlman, Michael Figora, Todd Fletcher, and Brian Scott who made the instrumentation of scintillator tile wrapping machine possible.

I would like to thank NIU physics faculty and staff. I would also like to thank my fellow graduate students Prudhvi Raj Varma Chintalapati, Christina Sarosiek, Brendan Leung, Aakaash Narayanan, Osama Mohsen, Ben Simons, Wei Hou Tan, Austin Dick, Andrew Best, Prudhvi Nikhil Bhattiprolu, Will Baker, Kevin Hamilton, Jeremiah Mitchell, and Sebastian Szustkowski for being great friends.

Lastly, I would like to thank my family for their constant support and love throughout my time away from home.

DEDICATION

I dedicate this dissertation to my mother, *Harsharan Kaur*; my father, *Lakhmir Singh*;
and my brother; *Gurpreet Singh* for their love and support.

TABLE OF CONTENTS

	Page
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF APPENDICES	xiii
Chapter	
1 INTRODUCTION	1
1.1 Standard model	2
1.1.1 Quantum Electrodynamics (QED)	4
1.1.2 Quantum Chromodynamics (QCD)	5
1.1.3 Electroweak Theory	6
1.1.4 Electroweak Symmetry Breaking and Higgs Mechanism	8
1.2 Vector Boson Scattering	11
1.2.1 Motivation	11
1.2.2 Topology of VBS in Semileptonic ZV Final State	13
2 THE LHC AND CMS EXPERIMENT	16
2.1 The Large Hadron Collider	16
2.1.1 Integrated Luminosity	18
2.2 The CMS Detector	19
2.2.1 The CMS Coordinate System	22
2.2.2 The Superconducting Magnet	23
2.2.3 The Tracking System	24

Chapter		Page
2.2.4	The Electromagnetic Calorimeter	25
2.2.5	The Hadronic Calorimeter	27
2.2.6	Muon Detector	28
2.2.7	Trigger	30
3	EVENT SIMULATION AND RECONSTRUCTION	32
3.1	Track Reconstruction and Calorimeter Clustering	32
3.2	Reconstructed Particles	34
3.2.1	Muons	34
3.2.2	Electrons and Photons	35
3.2.3	Hadrons and Jets	36
3.2.3.1	N-Subjetiness and Deep Taggers	38
3.2.3.2	Softdrop Mass	40
3.2.4	Missing transverse momentum	41
3.3	Monte Carlo (MC) Simulation	42
3.3.1	Generators	42
3.3.2	Parton Shower and Hadronization	43
3.3.3	Tunes	44
4	VBS MEASUREMENT IN ZVJJ FINAL STATE	45
4.1	Dataset and Simulation	46
4.1.1	Data	46
4.1.2	MC Simulations	47
4.2	Event Selection	52
4.2.1	VBS Tagged Jets and V Jet Candidate	52

Chapter	Page
4.3 Control and Signal Region	53
4.3.1 MC Scale Factors	54
4.3.2 DY+Jets Normalization	55
4.4 Kinematics Distributions	57
4.4.1 Boosted ZV DY+Jets Control Region	58
4.4.2 Resolved ZV DY+Jets Control Region	63
4.5 Machine Learning Modeling	68
4.5.1 Algorithm: Gradient Boosted Decision Tree	68
4.5.2 Training and Results	69
4.5.3 MVA Score Inference	80
4.6 Measurement	81
4.6.1 Statistical and Systematic Uncertainties	82
4.6.2 Significance	83
4.6.3 Impact Plots	85
4.6.4 Postfit Plots	87
4.7 Results Discussion and Future Prospects	90
5 HIGH GRANULARITY CALORIMETER UPGRADE	91
5.1 Technical Design and Requirements	91
5.2 Scintillator Tiles and SiPMs	94
5.2.1 Scintillator Materials	95
5.2.2 SiPMs	95
5.2.3 Scintillator Tiles	97
5.3 Automated Wrapping of Scintillator Tiles	99

Chapter	Page
5.4 Signal-to-Noise Ratio	103
5.4.1 Formulation	103
5.4.2 Testbeam and SiPM Noise Inputs	104
5.4.3 Scenarios	105
5.5 Results and Conclusion	106
6 SUMMARY	110
REFERENCES	111
APPENDICES	119

LIST OF TABLES

Table	Page
4.1 Trigger paths used to select events in CMS collision data	47
4.2 List of MC samples for Signal and Background modeling	49
4.3 List of MC samples for Signal and Background modeling	50
4.4 List of MC samples for Signal and Background modeling	51
4.5 Normalization factors from 1D bin of p_T of Z	56
4.6 Combined dataset Training and Testing Statistics in Signal Region	70
4.7 Training Input Variable Ranking	77
4.8 TMVA configuration of trained models for Boosted and Resolved ZV	77
4.9 Expected Signal Significance	84
4.10 Prefit and postfit yields of MC processes	88
4.11 Prefit and postfit yields of MC processes	89
5.1 HGCAL scenarios comparison	108

LIST OF FIGURES

Figure	Page
1.1 Standard model list of matter and interaction particles	3
1.2 3D representation of Higgs potential	10
1.3 The cross-sections for longitudinal polarized VBS involving three and four-boson coupling with and without Higgs coupling included.	12
1.4 Tree level Feynman diagram of VV VBS	13
1.5 Tree level processes that can happen in scattering represented by blob in VV VBS	13
1.6 Tree level Feynman diagram of ZV VBS process at LHC	14
2.1 A schematic of the CERN accelerator complex	17
2.2 Integrated delivered and recorded luminosity for 2015–2018 proton-proton collisions	18
2.3 The CMS detector cutaway view	20
2.4 Schematic view of the CMS detector	21
2.5 The picture of the CMS detector central part when lowered in underground cavern with superconducting magnet and iron yoke	23
2.6 The CMS pixel upgrade	25
2.7 The CMS ECAL schematic layout	26
2.8 The ECAL endcap quadrant assembled view	26
2.9 The first half of the barrel HCAL inserted into the superconducting solenoid (April 2006)	28
2.10 The HCAL depth segmentation after phase 1 upgrade	29

Figure	Page
2.11 The quadrant view of CMS subdetectors layout, and the coverage of the muon detector DTs, CSCs, and RPCs highlighted	30
3.1 Comparison of τ_{21} and τ_{32} shapes for signal and background in AK8 jets	39
3.2 The network architecture of DeepAK8	40
4.1 Normalization factors from 1D bin of p_T of Z and VBS trailing jet	56
4.2 DY+Jets Control Region: Leading electron kinematics in Boosted ZV Channel	58
4.3 DY+Jets Control Region: Leading muon kinematics in Boosted ZV Channel	59
4.4 DY+Jets Control Region: Hadronic boson kinematics in Boosted ZV Channel	60
4.5 DY+Jets Control Region: Leading VBS tagged jet kinematics in Boosted ZV Channel	61
4.6 DY+Jets Control Region: Trailing VBS tagged jet kinematics in Boosted ZV Channel	62
4.7 DY+Jets Control Region: Leading electron kinematics in Resolved ZV Channel	63
4.8 DY+Jets Control Region: Leading muon kinematics in Resolved ZV Channel	64
4.9 DY+Jets Control Region: Hadronic boson leading and trailing jet p_T in Resolved ZV Channel	65
4.10 DY+Jets Control Region: Leading VBS tagged jet kinematics in Resolved ZV Channel	66
4.11 DY+Jets Control Region: Trailing VBS tagged jet kinematics in Resolved ZV Channel	67
4.12 Decision Tree	69
4.13 Inputs Variables for training Boosted ZV BDT Classifier	71
4.14 Inputs Variables for training Boosted ZV BDT Classifier in signal region. . .	72
4.15 Inputs Variables for training Resolved ZV BDT Classifier	73

Figure	Page
4.16 Inputs Variables for training Resolved ZV BDT Classifier in signal region	74
4.17 Inputs Variables for training Resolved ZV BDT Classifier	75
4.18 Inputs Variables for training Resolved ZV BDT Classifier in signal region	76
4.19 Correlation Matrix for Signal and Background	78
4.20 MVA Score ROC Curve	79
4.21 MVA Score in Control Region for Boosted ZV Channel	80
4.22 MVA Score in Control Region for Resolved ZV Channel	80
4.23 MVA Score in Signal Region for Boosted ZV Channel	81
4.24 MVA Score in Signal Region for Resolved ZV Channel	81
4.25 Impact Plots of nuisance parameters from 1 to 30	85
4.26 Impact Plots of nuisance parameters from 31 to 60	86
4.27 MVA Score postfit in Signal Region for Boosted ZV Channel	87
4.28 MVA Score postfit in Signal Region for Resolved ZV Channel	87
5.1 Overview of HGCAL position and quadrant view of CE	93
5.2 CE-H mixed layer 22	94
5.3 SiPM	96
5.4 Scintillator tile with dimple	97
5.5 ESR wrapper cut for tile size R18–19	98
5.6 R18–19 scintillator tile wrapped in ESR	98
5.7 Automated scintillator tile wrapper overview	99
5.8 Tile folding station and ESR magazine	100
5.9 Sticker and tile dispenser	102
5.10 Tile folding station	102
5.11 Scintillator performance with various active area size of SiPM	107

Figure	Page
5.12 HGCAL scenarios	108
B.1 DY+Jets Control Region: Trailing electron kinematics in Boosted ZV Channel	122
B.2 DY+Jets Control Region: Trailing muon kinematics in Boosted ZV Channel	123
B.3 DY+Jets Control Region: Z boson kinematics in Boosted ZV Channel	124
B.4 DY+Jets Control Region: Trailing electron kinematics in Resolved ZV Channel	125
B.5 DY+Jets Control Region: Trailing muon kinematics in Resolved ZV Channel	126
B.6 DY+Jets Control Region: Z boson kinematics in Resolved ZV Channel	127
B.7 DY+Jets Control Region: V boson leading and trailing jet η in Resolved ZV Channel	128
B.8 Impact Plots of nuisance parameters from 61 to 90.	129
B.9 Impact Plots of nuisance parameters from 91 to 110.	130

LIST OF APPENDICES

Appendix	Page
A ANALYSIS CODE	119
B ADDITIONAL FIGURES	121
B.1 Control Region DY+Jets Boosted ZV Channel	122
B.2 Control Region DY+Jets Resolved ZV Channel	125
B.3 Impact Plots	129

CHAPTER 1

INTRODUCTION

It doesn't matter how beautiful your theory is, it doesn't matter how smart you are.

If it doesn't agree with experiment, it's wrong. In that simple statement is the key to science.

— Richard P. Feynman

“Particle Physics” is the branch of physics which deals with the fundamental particles and the interactions between them. Fundamental particles are the subatomic particles which at our present level of understanding are not made of other particles. There are two types of fundamental particles “matter” and “interaction” particles. As the name suggests matter particles are the fundamental constituent of matter and the interactions among them is governed by how they exchange interaction particles. Standard Model (SM) of particle physics is the theory that classifies these fundamental particles and describes three out of four fundamental interaction forces; electromagnetic, weak, and strong.

This chapter briefly introduces the theory of SM, Higgs mechanism, spontaneous electroweak symmetry breaking (EWSB), vector boson scattering (VBS), and the motivation for the search of VBS in semileptonic decay channel ZV with leptonic decay of Z , and hadronic decay of V (W/Z) to pair of quarks.

1.1 Standard model

In SM, the matter particles are fermions, and the interaction particles are bosons. SM also includes anti-fermions, which are fermions with equal mass but opposite sign of charge. Figure 1.1 lists mass, electric charge, and spin of fermions and bosons in the SM.

Fermions obey Fermi-Dirac statistics and have half integer spin. They can be further divided into leptons which have integral electric charge, and quarks which have fractional electric charge. There are three generations of quarks and leptons discovered to date, each generation only differing in mass. In addition to the electric charge, quarks also have three types of “color” charge (red, green and blue). Quarks cannot be isolated because of “color confinement”, which requires net color charge to be zero for a free particle, for this reason we can only have certain composition of quarks. Baryons (proton, neutrons, etc.) are made up of three quarks each with different color charge, and mesons (pions, kaons, etc.) are made of two quarks with color and anti-color charge.

Bosons obey Bose-Einstein statistics and have integral spin. They are described by a local gauge theory and are also called gauge boson. Photons are the interaction particle of electromagnetic force, they are massless and only interact with charged particles. Gluons are the mediator of strong force between the quarks, they are massless and carry color charge. W^\pm and Z are the vector bosons and mediator of weak force, unlike photons and gluons they are massive. W^+ and W^- are antiparticles of each other, and Z is its own antiparticle. The last gauge boson, Higgs is a massive scalar boson with zero spin, zero electric, and no color charge. Higgs boson is not a force carrier, but rather explains why only some particles have mass.

The SM is built in the framework of quantum field theory (QFT), in which particles are excitation of the fields and interactions arise from local gauge invariance. The SM is a

$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge theory, $U(1)_Y$, $SU(2)_L$, $SU(3)_C$ are the gauge symmetries of quantum electrodynamics (QED), weak interaction and quantum chromodynamics (QCD) respectively, where the indices stands for “hypercharge” (Y), “left-handed” (L) and “color” (C).

Standard Model of Elementary Particles

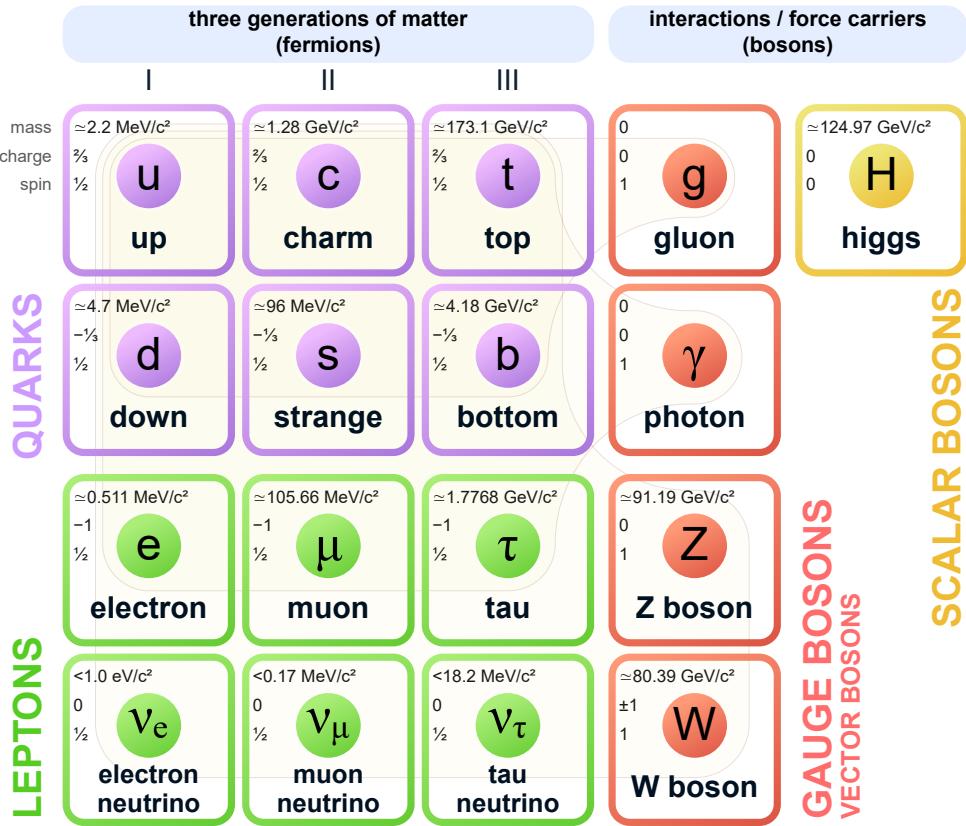


Figure 1.1: Standard model list of matter and interaction particles [1]

1.1.1 Quantum Electrodynamics (QED)

QED is a quantum field theory of electrodynamics, it describes the interaction of photons to the charged fermions. The QED is local gauge invariant and symmetric with $U(1)_Q$ symmetry group, defined as,

$$U(1)_Q = \exp(iQ\theta(x)) \quad (1.1)$$

where $\theta(x)$ is any spacetime function also called gauge parameter, and Q is the coupling constant of photon field to the fermions which is equivalent to the charge of fermion.

Under this transformation, fermion spinor $\psi(x)$ and four-potential A_μ electromagnetic tensor will transform as,

$$\psi(x) \rightarrow U(1)_Q \psi(x) \quad (1.2)$$

$$A_\mu \rightarrow A_\mu - \frac{1}{e} \partial_\mu \theta \quad (1.3)$$

The general Lagrangian of QED for fermions and their interaction with photon field is given by,

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.4)$$

where m is the mass of fermion, D_μ is the covariant derivative, and $F_{\mu\nu}$ is the electromagnetic field tensor defined as,

$$D_\mu = \partial_\mu + iQA_\mu \quad (1.5)$$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (1.6)$$

1.1.2 Quantum Chromodynamics (QCD)

The strong interactions are represented by $SU(3)_C$ gauge group, invariant under transformations of color charge degree of freedom, based on Yang-Mills theory [2]. Since “electrodynamics” is the theory of electric charge, this theory of color (*chromo* in Greek) charge is called “chromodynamics”, hence the name quantum chromodynamics (QCD).

A quark spinor in initial state can be represented as,

$$\psi = \begin{pmatrix} \psi_{red} \\ \psi_{blue} \\ \psi_{green} \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} \quad (1.7)$$

$SU(3)_C$ is an exact symmetry, it means the difference between colors cannot be measured experimentally, thus the color labels in quark spinor are arbitrary. $SU(3)_C$ transformation is defined as,

$$SU(3)_C = \exp \left(i\theta^a(x) \frac{\lambda^a}{2} \right) \quad (1.8)$$

where λ^a for $a = 1, \dots, 8$, are the Gell-Mann matrices, and $\theta^a(x)$ are any gauge parameters. These eight generators of symmetry correspond to eight gauge vector boson gluons.

Similar to QED, the covariant derivative for QCD can be formed as,

$$D_\mu = \partial_\mu + ig_s \frac{\lambda^a}{2} G_\mu^a \quad (1.9)$$

where g_s is the coupling constant of gluon to the quarks, and G_μ^a are the eight gauge fields corresponding to gluons.

The corresponding field strength tensor in QCD can be formed as,

$$F_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_\mu^b G_\nu^c \quad (1.10)$$

where f^{abc} are the structure constants of $SU(3)_C$ which satisfy $[\lambda^a, \lambda^b] = i f^{abc} \lambda^c$ relation.

The full Lagrangian for QCD can now be constructed as,

$$\mathcal{L}_{QCD} = \bar{\psi}^i (i\gamma^\mu D_\mu{}^{ij} - m\delta^{ij}) \psi^j - \frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu} \quad (1.11)$$

for mass m , indices i and j runs from 1 to 3.

The main difference of gluon field with respect to photon field is the presence of third term in field strength tensor which allows triplet and quartic self coupling of gluons.

1.1.3 Electroweak Theory

Weak interaction occurs for all the fermions via the exchange of massive vector bosons W^\pm , Z . Since the unification of electromagnetic and weak interaction into electroweak (EW) interaction by Glashow, Weinberg, and Salam [3–5], the weak interaction is better understood in terms of EW theory.

Weak interaction only couples to left-handed fermions and it is same whether the fermion is charged or not. The underlying gauge group of EW interaction is $SU(2)_L \otimes U(1)_Y$ and has

two transformations one for the left-handed doublet L and the right handed singlet fermions ψ_R which are defined as,

$$SU(2)_L \otimes U(1)_Y = \exp \left(i\theta^a(x) \frac{\sigma^a}{2} + i\theta(x) \frac{Y}{2} \right), \quad (\text{doublet}) \quad (1.12)$$

$$= \exp \left(i\theta(x) \frac{Y}{2} \right), \quad (\text{singlet}) \quad (1.13)$$

where Y is the hypercharge (linear combination of electric charge and weak isospin component), and σ^a for $a = 1, 2, 3$ are the Pauli spin matrices generator of $SU(2)$ symmetry. Left-handed fermion L doublets are,

$$L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau_L \end{pmatrix}, \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \quad (1.14)$$

and right-handed singlets are,

$$\psi_R = e_R, \mu_R, \tau_R, u_R, d_R, c_R, s_R, t_R, b_R \quad (1.15)$$

The covariant derivative of EW is then,

$$D_\mu L = \left(\partial_\mu + ig_w \frac{\sigma^a}{2} W_\mu^a + ig \frac{Y}{2} B_\mu \right) L \quad (1.16)$$

$$D_\mu \psi_R = \left(\partial_\mu + ig \frac{Y}{2} B_\mu \right) \psi_R \quad (1.17)$$

where W_μ^a and B_μ are the gauge fields. The EW Lagrangian can now be written as,

$$\mathcal{L}_{EW} = i\bar{L}\gamma^\mu D_\mu L + i\bar{\psi}_R \gamma^\mu D_\mu \psi_R - \frac{1}{4} W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \quad (1.18)$$

where $B_{\mu\nu}$ and $W_{\mu\nu}^a$ are fields strength, defined as,

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (1.19)$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g_w \epsilon^{abc} W_\mu^b W_\nu^c \quad (1.20)$$

the linear combination of B_μ and W_μ gauge field, with a weak mixing angle θ_w gives 4 vectors boson W^+ , W^- , Z , and γ of SM,

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp W_\mu^2) \quad (1.21)$$

$$Z_\mu = \cos \theta_w W_\mu^3 - \sin \theta_w B_\mu \quad (1.22)$$

$$A_\mu = \sin \theta_w W_\mu^3 + \cos \theta_w B_\mu \quad (1.23)$$

$$\tan \theta_w = g/g_w \quad (1.24)$$

Similar to QCD, the presence of third term in field strength tensor allows the self triple (WWZ , $WW\gamma$) and quartic ($WWWW$, $WWZZ$, $WWZ\gamma$, $WW\gamma\gamma$) couplings.

1.1.4 Electroweak Symmetry Breaking and Higgs Mechanism

Spontaneous symmetry breaking is the phenomena which explains why the ground state is not invariant under the symmetry of the Lagrangian. “Spontaneous” means that the symmetry breaking is not done by external agent but rather by Lagrangian itself in ground state.

The EW theory unifies weak interaction and QED but the gauge bosons in EW theory are all massless, if we were to add mass terms like $-m^2 W_\mu W^\mu$ by hand, it will no longer be gauge invariant. The solution to this, without breaking gauge invariance, is spontaneous

symmetry breaking, but this requires addition of new scalar field called Higgs field via Brout–Englert–Higgs mechanism (BEH) [6, 7], and this symmetry breaking is known as electroweak symmetry breaking (EWSB).

BEH introduces a complex scalar field as $SU(2)_L$ doublet with non-zero vacuum expectation value (VEV),

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^3 \end{pmatrix} \quad (1.25)$$

and BEH Lagrangian is,

$$\mathcal{L}_{BEH} = |D_\mu \phi|^2 - V(\phi) \quad (1.26)$$

where D_μ is same as EW covariant derivate in Equation 1.16, and $V(\phi)$ is,

$$V(\phi) = \mu^2 |\phi|^2 + \lambda(|\phi|^2)^2 \quad (1.27)$$

the parameter λ is required to be positive, for $\mu^2 > 0$ the minima is at 0, which is not an interesting case, but for $\mu^2 < 0$ vacuum state energy is given by,

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \quad (1.28)$$

by the choice of non-zero VEV v , scalar field can be parameterized as,

$$v = \sqrt{\frac{-\mu^2}{\lambda}} \quad (1.29)$$

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ h(x) + v \end{pmatrix} \quad (1.30)$$

where $h(x)$ is the Higgs field. By convention the v is selected to be real. This gives $\phi^1 = \phi^2 = \phi^3 = 0$ and $\phi^4 = v$ for the vacuum state. In making this selection the symmetry is spontaneously broken,

$$SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{EM} \quad (1.31)$$

Visually the Higgs potential is shown in Figure 1.2. The ball position at the center represents unbroken symmetry, and at the minima represents spontaneous broken symmetry in the ground state of potential.

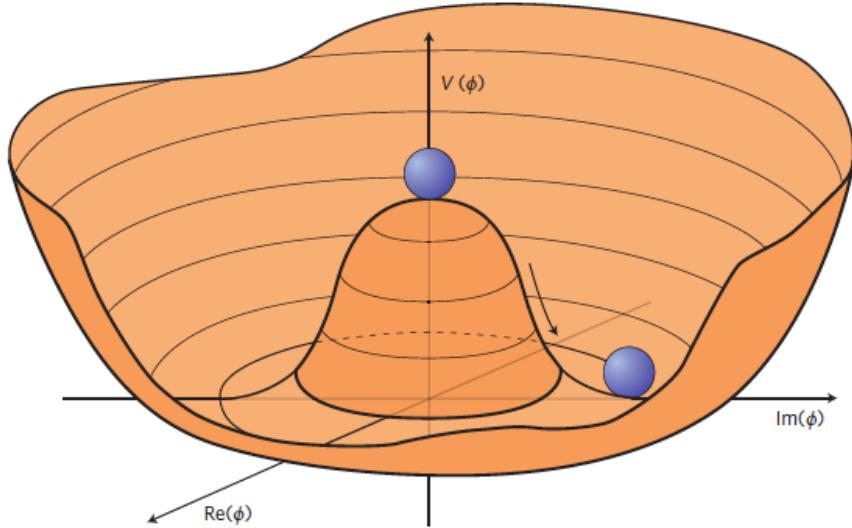


Figure 1.2: 3D representation Higgs potential [8].

After the symmetry is broken, the BEH Lagrangian contains the following mass terms,

$$m_W^2 W_\mu^+ W^{-\mu}, \quad m^2 Z Z_\mu Z^\mu, \quad m^2 h^2 \quad (1.32)$$

and gauge fields,

$$-\frac{1}{4} A_{\mu\nu} A^{\mu\nu}, \quad -\frac{1}{4} W_{\mu\nu}^+ W^{-\mu\nu}, \quad -\frac{1}{4} Z_{\mu\nu} Z^{\mu\nu}, \quad -\frac{1}{4} (\partial_\mu h)(\partial^\mu h) \quad (1.33)$$

thus explaining existence of three massive vector boson (W^\pm , Z), one massless vector boson (γ), and one massive scalar boson Higgs (H).

With experimentally measured value of VEV v approximately 246 GeV, the masses of bosons can be written in terms of v as,

$$m_A = 0, \quad m_W = \frac{g_W v}{2}, \quad m_Z = \frac{\sqrt{g_W^2 + g^2} v}{2}, \quad m_H = \sqrt{2\lambda} v \quad (1.34)$$

1.2 Vector Boson Scattering

The vector boson scattering (VBS) is a key process to experimentally study the nature of EWSB mechanism, because massive vector boson at high energy scattering can probe EW theory at TeV scale. This section describes the motivation behind studying VBS, and the topology of scattering of ZV channel studied in this dissertation.

1.2.1 Motivation

A massless spin-1 boson can exist in two transverse polarization as,

$$\varepsilon_\pm^\mu = \mp \frac{1}{\sqrt{2}} (0, 1, \pm i, 0) \quad (1.35)$$

and massive vector bosons can also exist in one longitudinal polarization,

$$\varepsilon_L^\mu = \frac{1}{m} (p_z, 0, 0, E) \quad (1.36)$$

This means the longitudinal polarized VBS cross-section will scale as E/m , whereas the scattering cross-section of transverse polarized boson remains constant. Figure 1.3 shows

the cross-section of longitudinal polarized VBS $V_L V_L \rightarrow V_L V_L$ from low to high energies. Perturbatively, the cross-section of longitudinal polarized VBS will scale with center of mass energy \sqrt{s} and eventually unitarity is violated at ≈ 1.2 TeV scale [9, 10]. Figure 1.3 also shows how the existence of light Higgs boson and inclusion of Higgs to vector boson coupling diagrams in longitudinal polarized VBS can restore unitarity, and any deviation from SM Higgs to vector bosons coupling will be visible in such a di-boson spectrum.

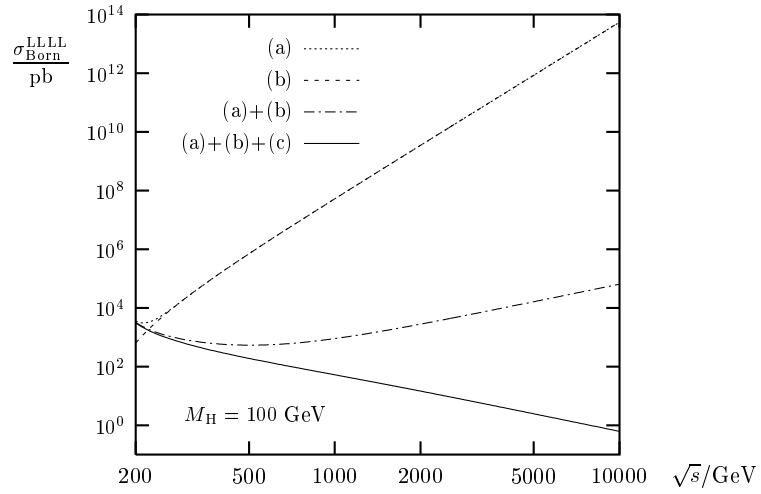


Figure 1.3: The cross-sections for longitudinal polarized VBS involving three and four-boson coupling with and without Higgs coupling diagram included. (a) is for the diagrams with three-boson coupling, (b) is for diagrams with four-boson coupling, and (c) is for diagrams with Higgs to vector boson coupling. [11]

Thus VBS measurement serves the two-fold purpose of doing precise measurements of the Higgs coupling to vector bosons at higher energy scales, and providing a framework to test for anomalous quartic gauge couplings in the SM.

1.2.2 Topology of VBS in Semileptonic ZV Final State

As discussed in previous section, VBS in SM can proceed via exchange of vector boson, Higgs boson, or quartic gauge coupling of four vector bosons. Figure 1.4, 1.5 shows tree level Feynman diagrams for VBS process.

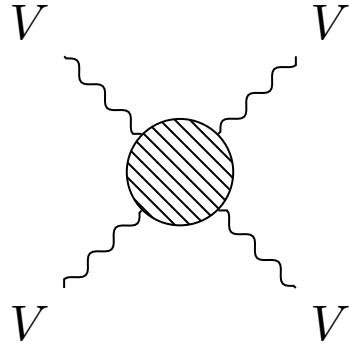


Figure 1.4: Tree level Feynman diagram of VV VBS, where V can be W^\pm , Z , or γ

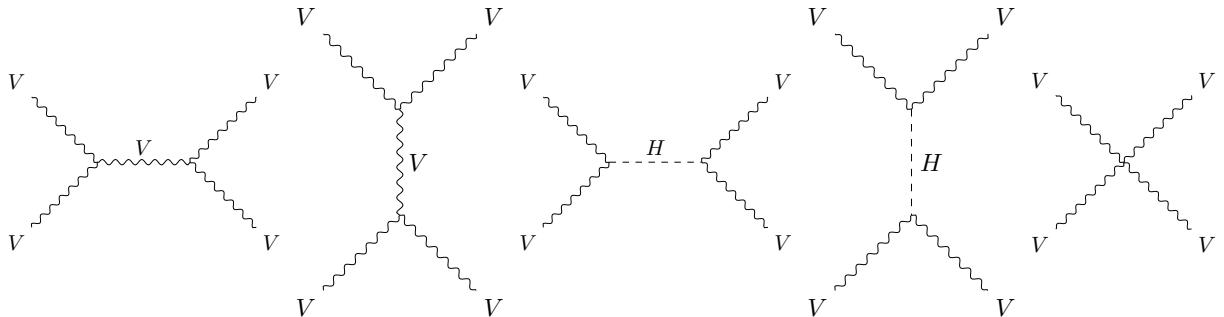


Figure 1.5: Tree level processes that can happen in scattering represented by blob in VV VBS, starting from s and t-channel exchange of vector boson, Higgs boson, and the last one is quartic coupling of vector bosons.

In proton-proton collisions, the actual interaction happens with the constituent quarks. For VBS to occur, the incoming (colliding) quarks have to radiate vector boson, then the scattering process between those vector boson can proceed via exchange of vector boson,

Higgs boson, or quartic coupling. The tree level Feynman diagram of a ZV VBS process in proton-proton collision is shown in Figure 1.6.

The outgoing quarks are a key signature of VBS in hadron collider experiments because they will have large pseudorapidity¹ difference between them, corresponding to a large invariant mass of outgoing quark pair. These outgoing quarks are first identified and tagged to filter out most of the QCD background.

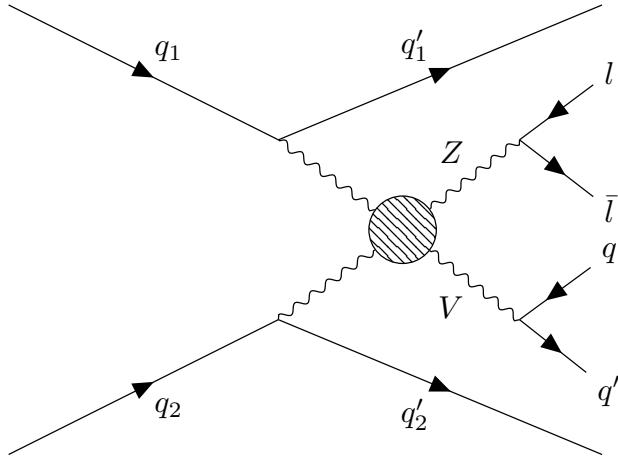


Figure 1.6: Tree level Feynman diagram of ZV VBS process at LHC. The diagram shows the production of two vector bosons being radiated from incoming quarks, in final state after scattering (blob), Z decays to pair of leptons, V (W/Z) decays to pair of quarks, and plus two outgoing quarks.

The type of leptonically decaying vector boson can be determined from the number of leptons in final state i.e. whether it was W (1 lepton + 1 neutrino) or Z (2 leptons). But given the hadronic activity in the proton-proton collisions, it is challenging to determine the type of hadronically decaying vector boson and is generally denoted by V .

Depending upon the decay channel of vector bosons, VBS can be categorized into fully leptonic or hadronic, or semileptonic channels. The fully leptonic channel has the cleanest signal, but the signal yield is also lower. The fully hadronic channel has the highest signal

¹defined as $\eta = -\ln[\tan \theta/2]$, where θ is angle between the z -axis and the beam-axis.

yield but also have large background because of hadronic activity in the final state. The semileptonic final state sits in-between these two channels, with cleaner boson reconstruction from leptons, and larger yields from hadronically decaying boson.

Out of two channels in the semileptonic, WV and ZV , WV has been studied and published by the Compact Muon Solenoid (CMS) [12]. Anomalous quartic gauge coupling limits derived using semileptonic channels with 2016 data of the CMS have also been published [13].

This analysis looks for the VBS signature with ZV in final state with the Z decaying to two opposite sign same flavor (OSSF) leptons, and V decaying to a pair of quarks in association with two outgoing quarks. Chapter 2 describes the proton-proton collision data used in this analysis and how various subdetectors of the CMS experiment work to capture the event information. Then in Chapter 3 the particle reconstruction and event simulations are discussed. In the Chapter 4 the object and events selection used are listed, the signal vs background classifier developed is described, and finally the measurement of signal significance is discussed.

CHAPTER 2

THE LHC AND CMS EXPERIMENT

The physics analysis is carried out using data collected by the CMS experiment at European Center for Nuclear Research (CERN) Large Hadron Collider (LHC) accelerator. This chapter provides overview of the LHC and details of the CMS experiment and its sub-detectors for particle tracking and calorimetry.

2.1 The Large Hadron Collider

The LHC is the largest proton-proton collider, located at CERN in Geneva, Switzerland. The main LHC ring is 27 km in circumference and 50 to 175 m underground. The LHC is built to collide protons at 14 TeV center-of-mass energy, LHC delivered proton-proton collisions at 7 and 8 TeV during run1 (2010–2012), and at 13 TeV center-of-mass energy during run2 (2015–2018) [14].

Figure 2.1 describes CERN accelerator complex. The protons are sourced by ionizing hydrogen atoms and then fed into a linear accelerator (LINAC). The LINAC accelerates the protons to 50 MeV and delivers them to the booster. Then the booster increases the energy of the protons to 1.4 GeV and feeds them to the proton syncrotron (PS) which further increases their energy to 25 GeV and starts bunching them together with bunches 25 ns apart. The proton bunches are passed through super proton synchrotron (SPS) which brings their energy to 450 GeV and bunches are then finally sent to main LHC clockwise and counterclockwise

rings where they are accelerated to final energy required which is 6.5 TeV for both bunches going clockwise and counterclockwise to obtain collisions at 13 TeV center-of-mass energy.

The proton-proton collisions occurs at four different locations where two general purpose detectors CMS and A Toroidal LHC Apparatus (ATLAS), and two specific purpose detector A Large Ion Collider Experiment (ALICE) and LHC-beauty (LHCb) are located.

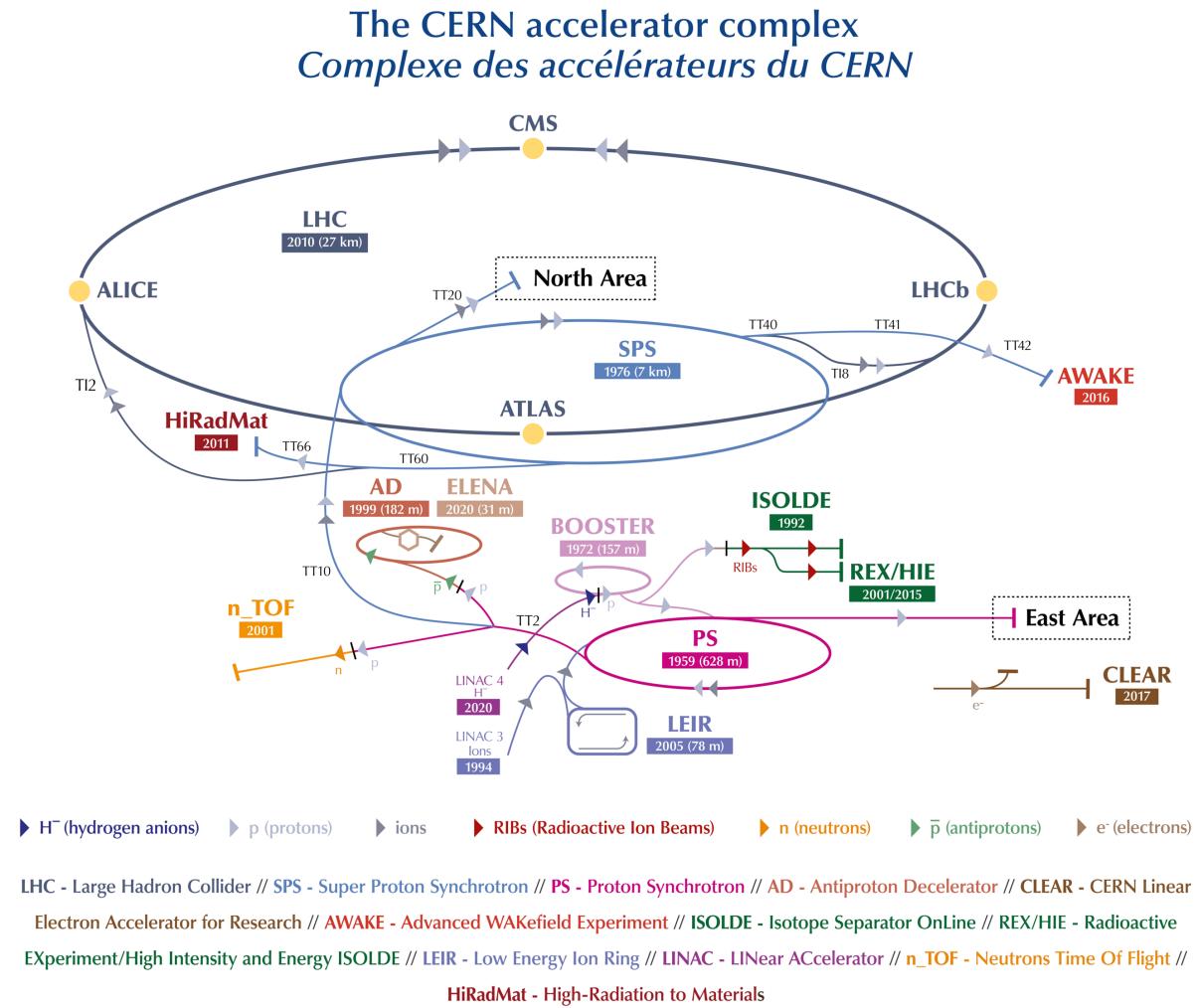


Figure 2.1: A schematic of the CERN accelerator complex [15]

2.1.1 Integrated Luminosity

Luminosity is a measure of how many events can happen per unit area and per unit time for a collision or scattering process, and it is defined as,

$$L = \frac{1}{\sigma} \frac{dN}{dt}, \quad L_{\text{int}} = \int L dt \quad (2.1)$$

where σ is cross-section of the process and L is the instantaneous luminosity ($\text{cm}^{-2} \text{s}^{-1}$), and L_{int} is the integrated luminosity usually expressed in inverse units of cross-section (pb^{-1} , fb^{-1} , etc.).

For proton-proton collision at LHC the nominal instantaneous luminosity is of the order of $10^{34} \text{ cm}^{-2} \text{s}^{-1}$, and integrated luminosity delivered and recorded by CMS during run2¹ operation is shown in Figure 2.2.

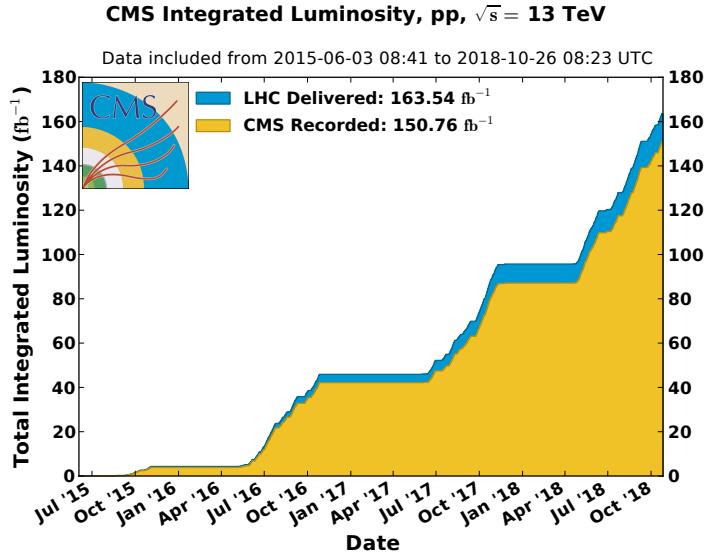


Figure 2.2: Integrated delivered and recorded luminosity for 2015–2018 proton-proton collisions [16]

¹2nd run period (2015–2018) of proton-proton collisions at LHC

For run2 standard physics analysis, luminosity recorded during 2016–2018 years is considered, and only runs certified as “golden” by CMS Luminosity physics object group (POG) are used for the analysis. The runs are certified as “golden” if all the subdetector, triggers and physics objects perform as expected. The total integrated luminosity for run2 standard physics is 137.19 fb^{-1} (35.92 fb^{-1} during 2016, 41.53 fb^{-1} during 2017, and 59.74 fb^{-1} during 2018 period) [17–19].

2.2 The CMS Detector

The CMS detector is a general purpose high energy particle physics detector. A cutaway view of the detector is shown in Figure 2.3. The detector is cylindrical with dimensions: 21 meters long, and 15 meters in diameter, with the whole detector weighing about 14000 tonnes. The detector is built in slices with central region called the “barrel”, and the two enclosing sides called “endcaps”. A superconducting solenoid generates magnetic field of 3.8 T inside and 2 T outside. To contain the magnetic field outside of solenoid and support structure of the detector, massive steel yokes are used.

Figure 2.4 shows a schematic view of the CMS detector, and how different particles leave signature in various sub-detectors. Neutral particles such as photons, neutrinos, and neutral hadrons will leave no track in the Silicon Tracker (ST) and are identified only by the energy deposited in the calorimeters or the missing energy in an event. Electrons are identified from the track in the ST and energy deposit in the Electromagnetic Calorimeter (ECAL). Since hadrons are heavier and they pass through the ECAL and deposit their most of energy in the Hadronic Calorimeter (HCAL), leaving only small fraction of energy in the ECAL. Since muons are minimum ionizing particles (MIPs), they pass through whole detector with very small fraction of the energy deposited in the ECAL and HCAL.

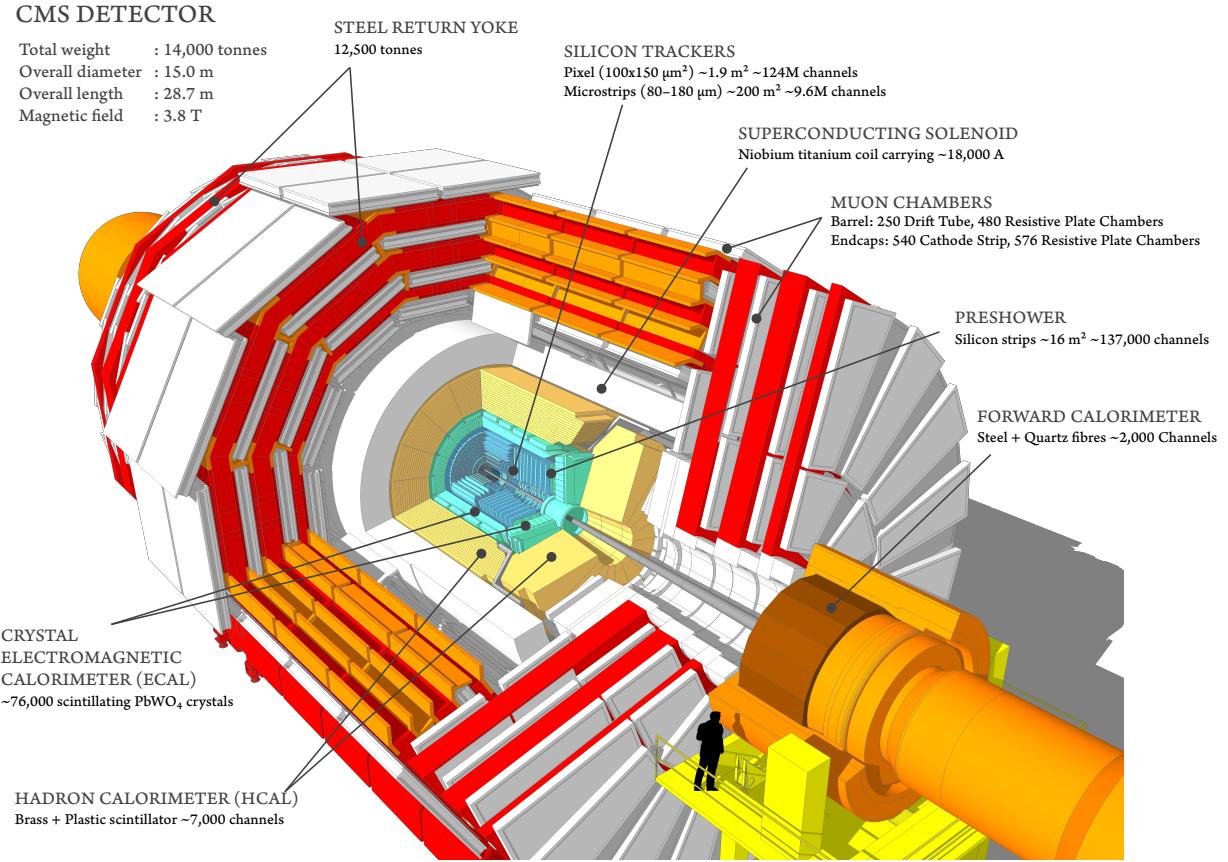


Figure 2.3: The CMS detector cutaway view [20]

This following section gives an overview of the subsystems of CMS detector. For detailed technical description refer to [21].

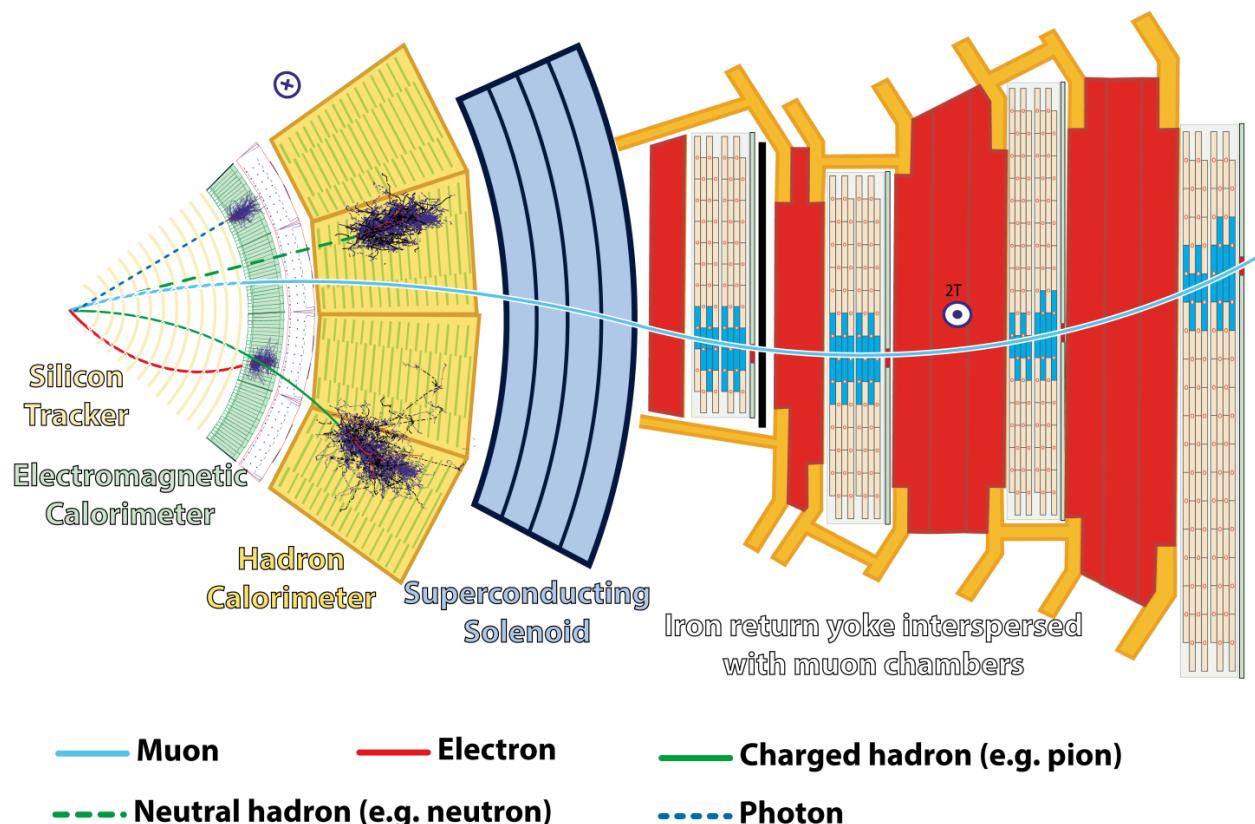


Figure 2.4: Schematic view of the CMS detector [22]

2.2.1 The CMS Coordinate System

CMS uses interaction point (IP) of the collisions as origin to define the right-handed coordinate system. The z -axis is along the beamline, the x -axis points toward the center of the LHC, and the y -axis points upwards toward the Earth's surface. The transverse plane $x - y$ is used to calculate most commonly used quantities like transverse momentum p_T and energy E_T .

To describe the direction of particles leaving the IP, azimuthal ϕ and polar θ angles are used. ϕ is measured around the beam axis, and θ is measured from the beam axis. In collider physics, pseudorapidity η (Lorentz invariant) is used to describe direction from beam pipe instead of θ as,

$$\eta = -\ln[\tan \theta/2] \quad (2.2)$$

and sometimes in terms of rapidity y as,

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad (2.3)$$

Particle kinematics can be completely described in terms of p_T , η , ϕ , and E_T or mass. The distance between the two particles in $\eta - \phi$ plane is described as ΔR ,

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (2.4)$$

2.2.2 The Superconducting Magnet

The superconducting magnet is the central part of the CMS detector, it is 12.5 meters long and 6.3 meters in diameter. The magnet is cooled to 4.5 K and a 20 kA current flows through it to generate 3.8 T of magnetic field with stored energy of 2.6 GJ.

Figure 2.5 shows the installation (in 2007) of the superconducting magnet with iron yoke in the underground cavern.

The key purpose of the magnet is to help determine the momentum and the sign of charged particles by bending them. The momentum resolution of the particles decreases with increase in their p_T . Constant magnetic field of 3.8 T gives momentum resolution of $\Delta p/p \approx 10\%$. This resolution is enough to determine unambiguously the sign of muons with momentum of $\approx 1 \text{ TeV}/c$.

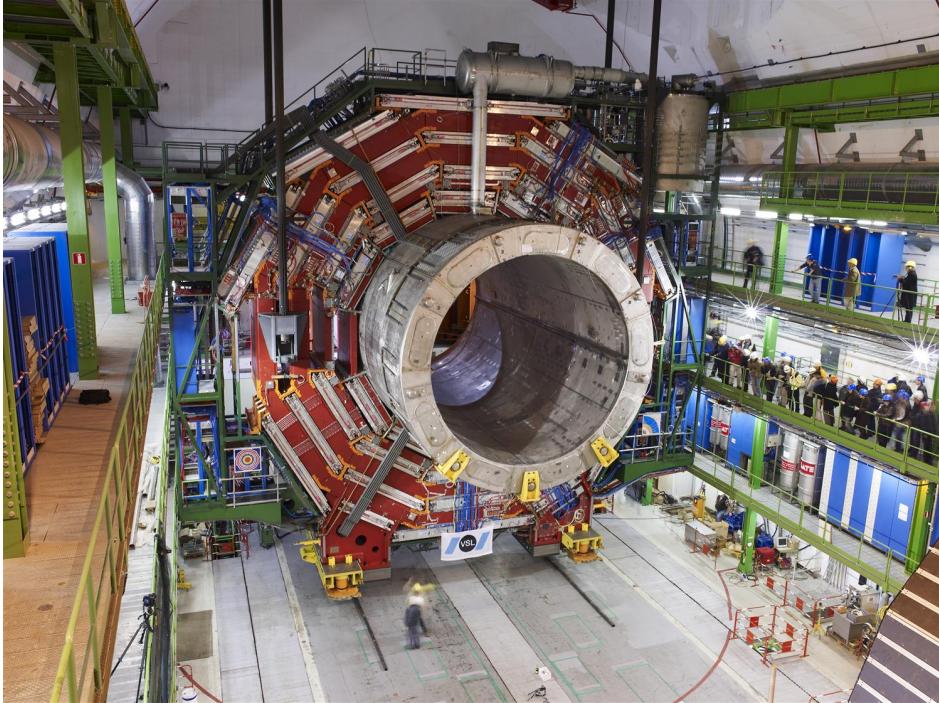


Figure 2.5: The picture of the CMS detector central part when lowered in underground cavern with superconducting magnet and iron yoke [23].

2.2.3 The Tracking System

The CMS tracking system ST is the innermost part of the detector, it is made up of pixel and strip detectors. The main goal of ST is to reconstruct the tracks of the charged particles with high precision in high pileup environment.

Silicon is most commonly used material for making tracking systems because of its semiconductor properties, and high radiation hardness which is essential for the innermost detector. When a p-n junction is built on silicon substrate it creates a depletion zone with no charge carriers at the junction, and whenever a charged particle passes through the depletion zone it creates an electron-hole pair. Under reverse bias this electron-hole generates an electrical signal. The CMS tracker consists of about 124 million channels of such junctions in the pixel detector and 10 million in the strip detector.

The pixel detector was upgraded in 2017 and the comparison of layers before and after the upgrade is shown in Figure 2.6. It is made up of four barrel layers and three endcaps, with nearest barrel layer being 3 cm away from the beamline for precise measurement of the IP. Because of the large number of pixel channels, the readout is done by Application-specific integrated circuits (ASICs).

The outermost part of ST detector is made up of silicon strips. It allows for large coverage by reducing number of readout channels. It has 10 layers in barrel region and 12 discs in endcap region. For better signal-to-noise ratio and radiation tolerance both pixel and strip detector are operated at -20 °C.

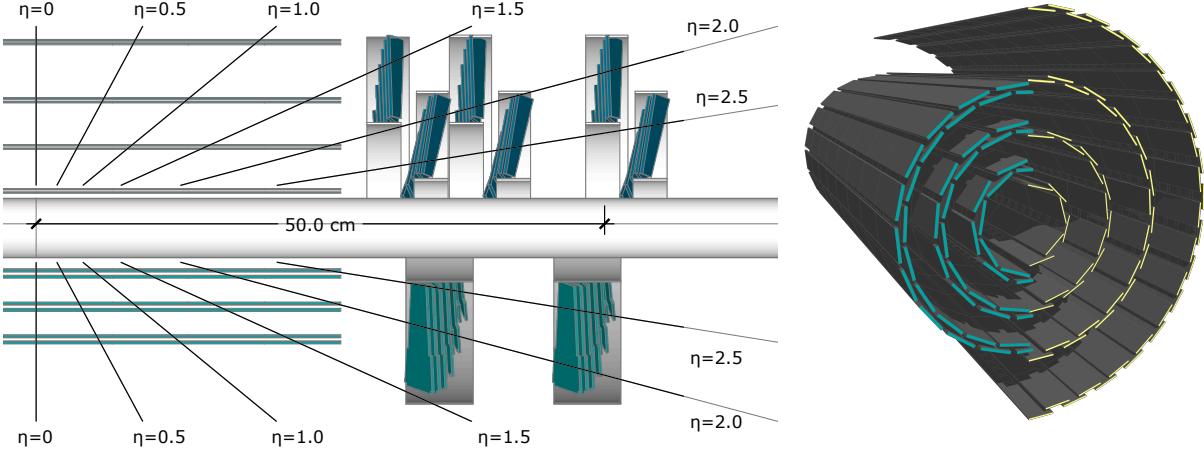


Figure 2.6: The CMS pixel upgrade. The left is cross sectional view of pixel detector layers before upgrade (bottom) and after Phase 1 upgrade (top). The right is view pixel barrel before upgrade (left) and after upgrade (right) [24].

2.2.4 The Electromagnetic Calorimeter

The ECAL active material is made of lead tungstate (PbWO_4) scintillating crystals and two layers silicon strip for preshower in front of the endcaps. The crystals in central barrel section are mounted in quasi-projective geometry pointing towards IP and cover $|\eta| < 1.48$. The two endcaps extend the coverage to $|\eta| = 3.0$. The schematic layout of ECAL is shown in Figure 2.7 and the picture of endcap quadrant when assembled is shown in Figure 2.8.

The main purpose of the ECAL is to determine energy and positions of electromagnetically interacting particles. Except electron and photons all other particles pass through ECAL crystals with only small fraction of energy signature in the crystals. When electron or photon interact with PbWO_4 it starts a process of electromagnetic shower which continues until the energy of the incident particle is below threshold, which is about 1 MeV.

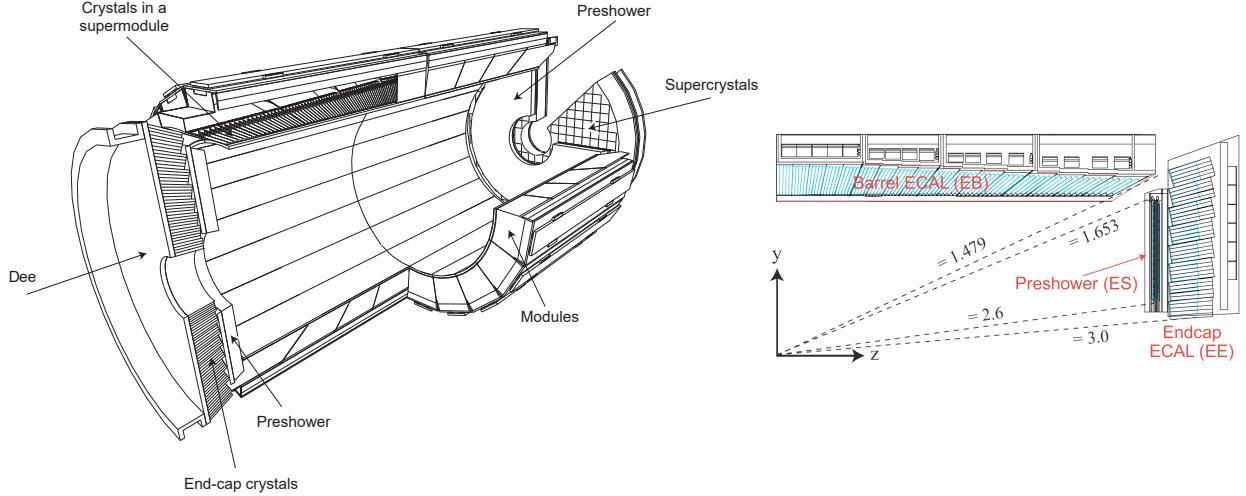


Figure 2.7: The CMS ECAL schematic layout. The left schematic shows the arrangement of superclusters in barrel and endcap (with preshower layers). On the right is the $y - z$ plane quarter view of ECAL layout [25].

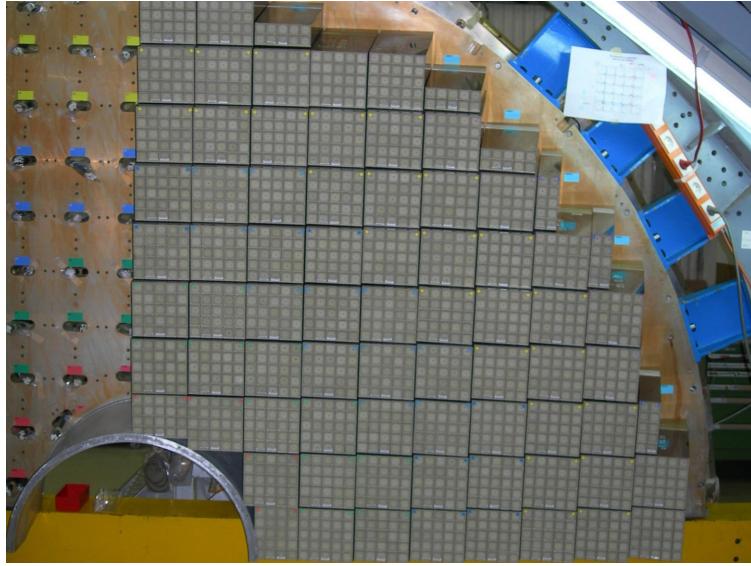


Figure 2.8: The ECAL endcap quadrant assembled view [26].

The fractional resolution of the ECAL energy measurements can be described as,

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (2.5)$$

where S is the stochastic term, N is related to the noise, and C is a constant offset. The energy resolution in the central region of ECAL barrel ($|\eta| < 0.8$) for electrons from Z decays was 2% and 2–5% elsewhere. The energy resolution for photons from Higgs decay varied from 1.1% to 2.6% in the barrel region and from 2.2% to 5% in the endcaps [27].

2.2.5 The Hadronic Calorimeter

Similar to ECAL, the purpose of HCAL is to shower hadrons, and measure their energy and position. The first half of barrel HCAL inserted into solenoid during installation is shown in Figure 2.9.

HCAL is made up of towers pointing towards IP where each tower is made up of sampling layers; with alternating layers of plastic scintillator and brass. Brass acts as absorber in the HCAL and causes hadrons to shower. Scintillators emits light when secondary particles from hadronic shower pass through them. The light output collected from the scintillator layers, combined with total absorption length of Brass layers gives the total energy deposited in the HCAL. In phase 1 upgrade the HCAL was upgraded to give energy deposit as a function of depth. The depth segmentation schematic is show in Figure 2.10 and the details of upgrade are in technical design report [28].

HCAL consists of barrel (HB) and two endcaps (HE) located inside solenoid. These two subsystems combined provides coverage of $|\eta| \leq 3.0$. There are two other subsystems of HCAL outside solenoid, a forward HCAL (HF) and outer barrel HCAL (HO). HO was added to ensure there is no leakage from the particles that make it past the solenoid. HF extends the coverage to $|\eta| \leq 5.0$. HF is based on Cherenkov radiation principle and uses quartz fiber as active material with steel absorbers.



Figure 2.9: The first half of the barrel HCAL inserted into the superconducting solenoid (April 2006) [29].

2.2.6 Muon Detector

The outermost subsystem in the CMS detector is the muon detector. Unlike electrons, muons are MIPs i.e. they do not lose much of their energy while passing through tracker, calorimeter and solenoid. Muon detector is built to identify, measure momentum and trigger the events with muons. Like other subsystems, the muon detector consists of a barrel and endcap detector. The schematic layout is highlighted in the Figure 2.11. As shown in the schematic, the muon detector consists of three subsystems drift tubes (DTs), cathode strip chambers (CSCs) and resistive plate chambers (RPCs).

The DTs are wire gas detectors filled with Argon and composed of many tube cells of about 4 cm. Muon passing through these tubes ionize Argon and the free electron is detected on the wire cathode. Each DT is about 2 meters by 2.5 meters in size, and there are four

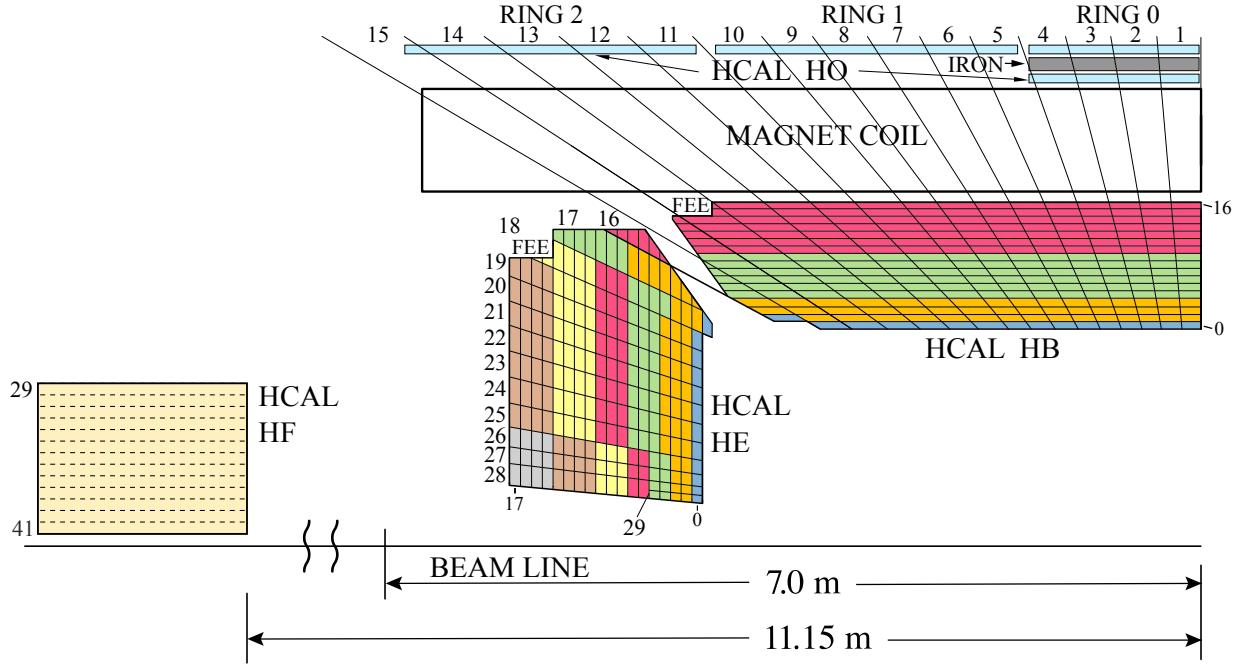


Figure 2.10: The HCAL depth segmentation after phase 1 upgrade [30].

layers of the DTs interleaved with the iron yoke, parallel to the beam pipe in barrel region. The drift time is the order of about 380 ns.

The CSCs are based on same principle as DTs and are made of multi-wire proportional chambers consisting of 6 anode planes interleaved with 7 cathode planes. They have time resolution better than 5 ns. The CSCs are used in endcap region, where radiation hardness is required, and non-uniform magnetic field does not affect the measurement.

The RPCs are made up of two oppositely charged high resistive parallel plates with a gas² volume between them. When a charged particles passes through, it ionizes the gas, and it creates an avalanche. The charge is collected by metallic readout strips. RPCs have poor position resolution but fast readout of the order of 1 ns, which is fast compared to DTs, this is the reason there are 1 or 2 RPCs attached to both DTs, and CSCs.

²mixture of 96.2% $C_2H_2F_4$, 3.5% Iso- C_4H_{10} , 0.3% SF_6

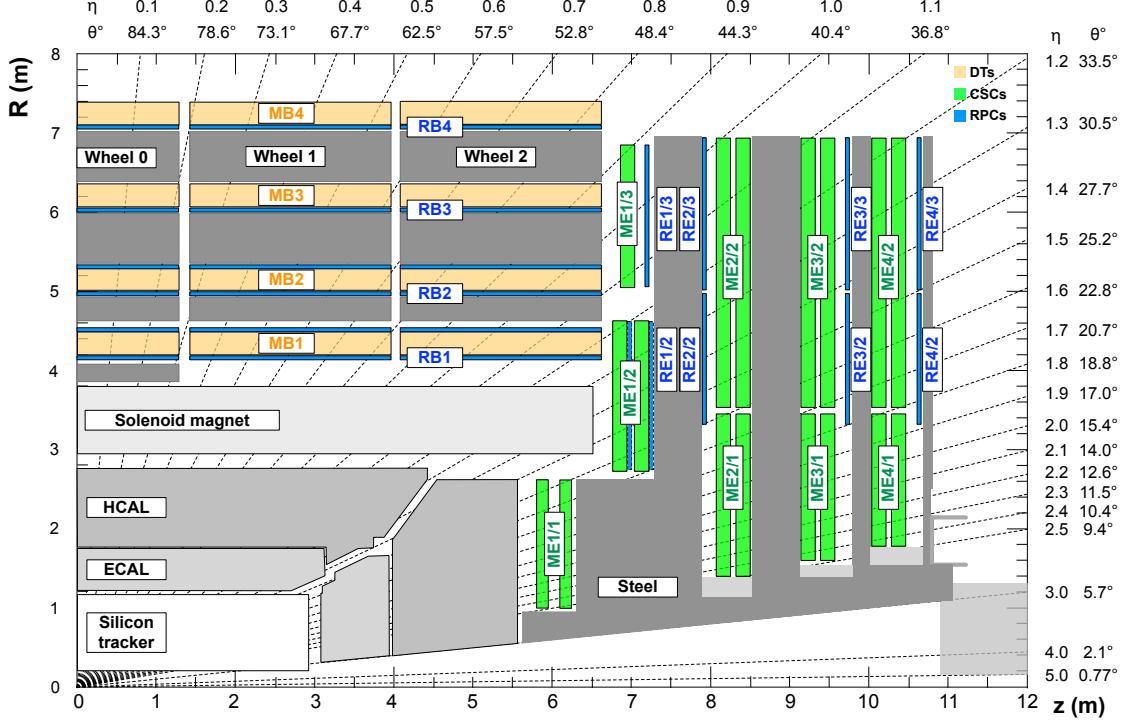


Figure 2.11: The quadrant view of CMS subdetectors layout, and the coverage of the muon detector DTs, CSCs, and RPCs highlighted [31].

2.2.7 Trigger

Since proton-proton collisions occurs every bunch crossing which are 25 ns apart; equivalent to a 40 MHz collisions rate. At this collisions rate, the data storage required will be enormous and CMS can only record up to 1000 events per second. Since most of events do not contain interesting physics events, they can be thrown away. To do this CMS has two tier trigger system Level-1 Trigger (L1T), and High Level Trigger (HLT).

The L1T is the first electronic processing system through which event information is processed before it is passed to second trigger system: HLT. The L1T is designed to make fast decisions in about $3.8\ \mu s$, and only uses ECAL, HCAL and muon system to make a decision regarding throwing or keeping an event. L1T cut downs the data rate from 40 MHz

to 100 kHz. The L1T electronics is placed next to the detector in underground cavern for fast transfer of data.

Events accepted by L1T are passed to the HLT. The HLT further reduces the data rate from 100 kHz to about 1 kHz using a computer farm with nearly 26000 cores. HLT uses all the available information from the event to make decision in about 300 ms. HLT is modular by design to allow the use of information from different systems to construct multiple paths called HLT paths, for example the single muon HLT path will save event with at least one muon passing the selection criteria set in HLT path. Events passing at least on HLT path are saved for offline physics analysis.

CHAPTER 3

EVENT SIMULATION AND RECONSTRUCTION

The proton-proton collision at the LHC produces a shower of particles. Before the event information can be easily used in an analysis, the data collected goes through an iterative process of reconstruction of particles produced in the collision. CMS uses a particle flow (PF) algorithm to reconstruct 4-vectors of muons, electrons, photons, hadrons, jets and missing transverse momentum [32].

To analyze the data collected and compare it with theoretical models, events are simulated using Monte Carlo (MC) event generators and are passed through detector simulation and PF reconstruction so that MC events can be treated in the same way as real events.

This chapter describes the basic ingredients for object reconstruction, PF candidates and MC event generators used in this analysis.

3.1 Track Reconstruction and Calorimeter Clustering

For complete particle reconstruction the two main ingredients are the track left by particle in the detector and the energy deposit in calorimeter. This section describes track reconstruction from the hits in the tracker and muon detector, and measurement of energy deposit from calorimeter clustering.

Track reconstruction requires reconstructed hits, and seed generation which are described in [33]. The track reconstruction is done using pattern recognition which is an iterative process and based on a combinatorial Kalman Filtering (KF) method [34]. Starting from

first hits pair from initial layers are used to estimate a track. Next hits information of subsequent layers are included to grow the tracks into possible trajectories. There can be multiple hits in each new layer, for this multiple trajectory candidates are created. All the trajectory candidates are grown in parallel to avoid bias and truncated at each layer to prevent exponential increase in number of candidates. Then finally the track is fitted to compute momentum and vertex information.

The main purpose of the calorimeter clustering is to determine the position and energy of the particle. A cluster in a calorimeter is a local group of energy deposits that are spatially consistent with an electromagnetic or hadronic shower. First the topological clusters are identified, a topological cluster is a contiguous region of energy deposits. Next seeds are identified in topological cluster with certain energy threshold and are local maxima within topological cluster. In ECAL seed is a crystal with energy threshold of 0.23 GeV and have highest energy deposit among the 8 neighboring crystals in a topological cluster. In HCAL seed is a tower with energy threshold of 0.08 GeV highest energy deposit among the 4 neighboring towers in a topological cluster.

Each seed is grown into energy cluster by gathering neighboring crystals (or towers) energy. For the case when there is one seed, there is only one iteration until all the neighboring crystals (or towers) are added. In case when there are more than one seeds, the energy from neighboring crystal (or towers) is shared. The fraction of energy shared depends on the energy and position of the clusters being reconstructed. Each iteration uses re-calculated energy and position of the clusters. Iteration continues until either the maximum iteration is reached, or clusters energy and positions values have converged.

3.2 Reconstructed Particles

After tracks and calorimeter clusters are formed, PF links this information from the detectors together to form objects as broadly discussed in Section 2.2 and shown in Figure 2.4. This section describes the properties of those reconstructed particle candidates.

3.2.1 Muons

Reconstructing muon with best precision is the key ingredient for many physics searches. Muons reconstruction and identification uses information from the tracker, calorimeters and muon detector. There are two types of reconstruction performed; “Global” and “Tracker” for muon candidates. Global muons are formed combining and refitting muon hits in the muon detector with compatible track from ST, and the tracker muons are formed by extrapolating tracks from ST to segments in muon detector.

Once the muon candidates are found, the kinematic properties (p_T, η, ϕ) are calculated from track fitting. Other properties such as distance from primary vertex (PV) dxy, dz , number of hits in the tracker and muon system, tracker based relative isolation (3.1) in a cone of $\Delta R = 0.3$, and PF based relative isolation (3.2) in a cone of $\Delta R = 0.4$ are stored for cleaning and isolating muons for physics analysis.

The tracker and PF based relative isolation are defined as,

$$\text{TkIso03} = \left(\sum p_T^{\text{Tracks (PV)}} \right) / (p_T^\mu) \quad (3.1)$$

$$\text{PFRellIso04} = \left(\sum p_T^{\text{CH (PV)}} + \min \left[0, \sum E_T^{\text{NH}} + \sum E_T^\gamma - 0.5 \sum p_T^{\text{CH (PU)}} \right] \right) / (p_T^\mu) \quad (3.2)$$

where “Tracks (PV)” refers to all the tracks in tracker and coming from PV, “CH (PV)” and “CH (PU)” refers to charged hadrons coming from PV and pile up (PU) respectively, “NH” refers to neutral hadrons, μ refers to muon, and γ refers to photon.

There are multiple source of muons whenever collision event happens, they can be real muons or hadrons which are misidentified as muons, these hadrons are able “punch” through HCAL and leaves hit in muon detector. The real muons of interest are called “prompt” muons and others are usually referred as “non-prompt”. Non-prompt muons can originate from decay of pions and kaons in flight usually identified with a “kink” in track or from heavy flavor decay of b or c-quarks which are identified with tracks not originating from PV. The prompt muons are the ones coming from decay of H, W, Z bosons and τ leptons, and have small impact parameter from PV, have hits in both tracker and muon detector, and are typically well isolated.

Muon identification for loose and tight muons follows CMS Muon Physics Object Group recommendation and are defined in Reference [35].

In addition to muons from collision events, there can be cosmic muons from pion decay in upper atmosphere. Cosmic muons are easily rejected since generally they are not in-time with collision and are far from interaction points.

3.2.2 Electrons and Photons

Since there is large amount of material in the tracker, electrons often emit bremsstrahlung photons when passing through tracker volume, and energetic photons can decay to e^-e^+ pair which complicates the tracking algorithm. The energy deposit of such electrons emitting bremsstrahlung will have large spread in ϕ direction because the magnetic field will bend electrons in ϕ whereas photons are unaffected. For this reason electron and photon

reconstruction are done together, and the Gaussian-sum filter (GSF) algorithm is used for electron track reconstruction which takes care of kinks in electrons track because of hard emission [36].

An electron is reconstructed when an ECAL cluster matches a GSF track, and a photon is reconstructed when an ECAL cluster with E_T more than 10 GeV is found and have no matching GSF track. To prevent electrons and photons from being misidentified as jets certain conditions are applied, for electron the number of GSF tracks matching with the ECAL cluster is limited to a maximum of two, and energy deposits in a cone of $\Delta R = 0.15$ in HCAL around the position of electrons and photons is required to be less than 10% of ECAL.

Similar to muons, after electron and photon reconstruction is done, their kinematics properties are calculated, and various other properties required for cut based and multivariate analysis (MVA) based identification are stored. The detailed description of electrons and photons identification technique and properties used in this dissertation can be found in Reference [37].

3.2.3 Hadrons and Jets

Quarks and gluons produced in a collision event are not detected directly, because of color confinement. They go through fragmentation and hadronization making a collimated spray of particles called “jets”. Charged hadrons are reconstructed when a HCAL cluster can be associated with one or more tracks, if the track association fails the cluster is reconstructed as neutral hadron.

Jet in CMS are reconstructed using the FASTJET package [38], which takes as input all of the PF candidates and the associated tracks. The clustering start with combining 4-vectors

of the two particles i and j , if the distance between them (d_{ij}) is lower than stopping distance (d_{iB}). Then particles i and j are removed from the input collection, and the clustering is continued with combined ij particle and the next candidate in the collection. The clustering stops when, $d_{ij} > d_{iB}$ and i is called a jet.

d_{ij} and d_{iB} are defined as,

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (3.3)$$

$$d_{iB} = p_{Ti}^{2p} \quad (3.4)$$

where p is the parameter for different clustering algorithms, R is the cone size, and ΔR_{ij} is distance between two particles in iteration.

Anti- k_T (AK) is the most used jet algorithm in CMS physics analyses. This corresponds to $p = -1$, which means the hard particles will be clustered first in this clustering algorithm. The cone size used for standard jets (AK4) is $R = 0.4$, and for large jets (AK8) often called “fatjet” is $R = 0.8$.

To mitigate the effect of PU contamination in jets two most commonly used techniques are charged hadron subtraction (CHS) and Pileup Per Particle Identification (PUPPI) [39]. CHS as the name suggests removes all the PF particles in the jet clustering which are originating from PU vertices, and it is a standard technique for AK4 jets. PUPPI rescales 4-vectors of all the particles in a jet with a weight. The PUPPI weight is derived from a local metric which is built using information of charged hadrons whether they are originating from leading vertex or PU. The main limitation of CHS is that it only removes charged PU contribution, for larger jets it can be issue, since it is clustering larger number of particles and can have significant contribution from neutral hadrons, for this reason PUPPI technique is used for AK8 jets.

Jet energy corrections are applied to the reconstructed jets to correct for jet energy scale (JES) differences in MC and data events. Three levels of corrections are applied. The first correction that is basically for the removal of pileup and electronic noise and is applied to both MC and data events. The second correction applied only to MC events corrects for jet response as a function of p_T and η to the particle level jet response. Third correction applied only to collision data are derived from the differences in p_T and η response of jets in data and MC samples, using $\gamma + \text{jets}$, $Z + \text{jets}$, and dijet MC events [40].

To improve the jet selection and reject jets originating purely from PU, two methods are used in this dissertation, (a) jet identification based on multiplicities and energy fraction of particles contained in the jet, and (b) MVA based PU identification which uses jets shape variables to discriminate prompt jet from pileup jets. The details of PU mitigation and identification used in CMS are described in Reference [41].

After jet reconstruction is complete, in addition to calculating kinematic properties (p_T , η , ϕ and mass), various other properties such as b-quark tagging and quark-gluon likelihood are also calculated and stored.

3.2.3.1 N-Subjetiness and Deep Taggers

The origin of fatjets is usually when a heavy energetic particle such as W or Z bosons decays to a pair of quarks and the resulting jets overlap each other. This usually happens when decaying the particle is boosted in lab frame, and the fatjet is usually referred as boosted jet. To find and discriminate the fatjet of interest based on its substructure the two common technique used in CMS are N-Subjetiness [42] and “deep tagger” [43].

N-Subjetiness is defined as,

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\} \quad (3.5)$$

where k runs over constituent particles in a jet, $\Delta R_{J,k}$ is the distance between subjet J and k constituent, and d_0 is the normalization constant defined as,

$$d_0 = \sum_k p_{T,k} R_0 \quad (3.6)$$

where R_0 is the jet radius ($= 0.8$, for fatjet).

τ_N quantifies to what degree a jet can be regarded as made of N jets. The small values of the τ_N means a jet is more likely to have N or less subjets, and higher value means it will at least have $N + 1$ subjets. Rather than using τ_N alone, ratio of different τ_N variables is preferred for rejecting QCD jets. Figure 3.1 shows distribution of $\tau_{21} = \tau_2/\tau_1$ and $\tau_{32} = \tau_3/\tau_2$ shapes for different MC samples. τ_{21} is used to discriminate fatjets with 2-prongs (W/Z/H) and τ_{32} with 3-prongs (t-quark) substructure against QCD jets.

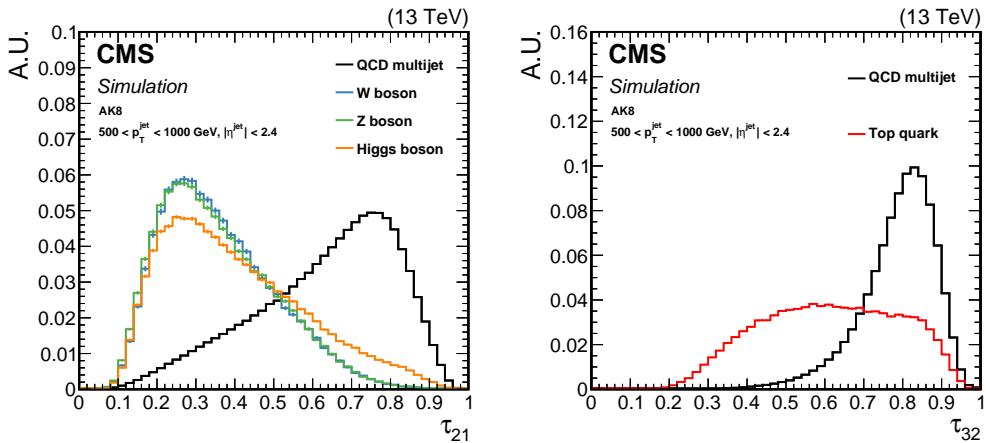


Figure 3.1: Comparison of τ_{21} and τ_{32} shapes for signal and background in AK8 jets. The left is τ_{21} distribution showing discrimination W/Z/H jets vs QCD jet, and the right is τ_{32} distribution for t-quark vs QCD jets [43].

Deep Tagger for AK8 is a machine learning (ML) based tagger developed to determine the origin of a fatjet. Figure 3.2 describes the architecture of “DeepAK8” tagging. These taggers are trained on particle level information from PF, and provide multi class tagging probabilities such as Z vs. QCD, W vs. QCD, etc. Since deep taggers utilize deep neural network with particle level information their classification power is higher than N-subjetiness and are worth the study for future analysis.

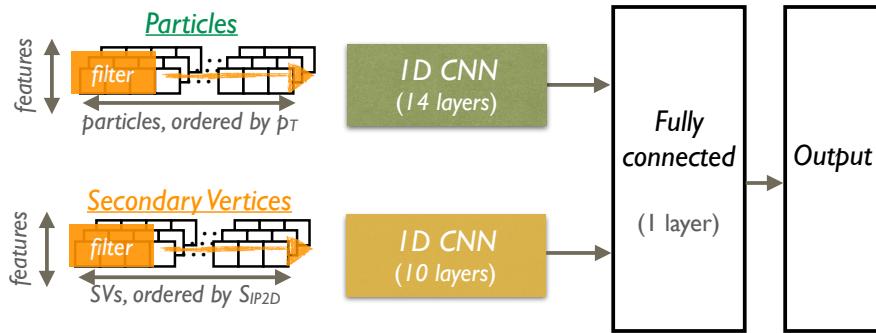


Figure 3.2: The network architecture of DeepAK8 [43]

3.2.3.2 Softdrop Mass

Fatjets can also have contamination coming from wide angle soft initial state radiation (ISR) and multiple hadron scattering, which affects the mass calculation of the jet, to remove such contamination and have better mass reconstruction, the “softdrop” mass algorithm [44] is used.

Softdrop is a declustering algorithm which removes the particle from the jet with radius R_0 ($= 0.8$, for fatjet), when the following condition between two particles is satisfied,

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut} \left(\frac{\Delta R_{12}}{R_0} \right)^\beta \quad (3.7)$$

where ΔR_{12} is the distance between the two particles, z_{cut} and β are the parameters for tuning softdrop declustering. For fatjets used in CMS, the softdrop algorithm is applied with $\beta = 0$ and $z_{cut} = 0.1$ which vetoes both soft and soft-collinear emissions in a jet.

3.2.4 Missing transverse momentum

Invisible particles like neutrinos cannot be detected at CMS directly. Kinematics of such particles can be determined using the law of conservation of total momentum. In case of proton-proton collision, the actual collision happens between quarks contained in proton which carry an undetermined fraction of proton momentum. For this reason kinematic determination of invisible particles is limited to transverse plane only.

After all the particles are reconstructed in an event, their p_T 's can be used to determine missing transverse momentum as,

$$\vec{p}_T^{miss} = - \sum \vec{p}_T \quad (3.8)$$

It's usually neutrinos which contributes to missing transverse momentum and they have very small mass, the missing transverse momentum is then equivalent to Missing Transverse Energy (MET), which is most often used term in physics analyses. Experimentally, noise in detector or beam halo of colliding beams can cause anomalously large missing transverse momentum. There are various algorithms available in CMS that can suppress such events.

3.3 Monte Carlo (MC) Simulation

In the proton-proton collision, determination of exact underlying process from reconstructed final state is not possible. To do measurements on the process of interest, we need to quantify how much is the signal (process of interest) and how much is the background (processes with same final state and similar kinematics). MC simulations are employed to generate underlying processes that can help in the modeling of signal and background process and compare them against the collision data.

Main tools and framework used at CMS for generating MC events are described in this section.

3.3.1 Generators

There are three main components to the event generator, parton distribution function (PDF), matrix element and parton shower. PDF encodes the probability of finding a patron with given fraction of energy at given proton momentum. Matrix element encodes the theoretical cross-section of the hard process.

Our signal $ZVjj$ is a pure EW process generated at leading order (LO) with MADGRAPH5_aMC@NLO and MADSPIN [45, 46]. At LO this process has six EW vertices, and zero QCD vertices. The generation is done in two steps with both vector boson generated on shell with MADGRAPH and decayed to leptons and quarks via MADSPIN. For example, to generate ZZ process, the syntax used is,

```
$ madgraph >
generate p p > z z j j QED=4 QCD=0
$ madspin >
decay z > l+ l-
decay z > j j
```

where p and j are defined as collection of quarks of gluons, z is the on-shell Z boson, and l^+ , l^- are the leptons consisting of electrons, muons and taus. The syntax indicates $\text{QED}=4$ $\text{QCD}=0$ implies that there are four EW vertices, since we are first generating Z on-shell, and zero QCD vertices. The last two EW vertices come from decay of Z bosons to fermions.

3.3.2 Parton Shower and Hadronization

After events are generated with the hard process of interest, they go through the process of parton shower and hadronization to simulate QCD bremsstrahlung radiation and make final state hadrons to preserve overall color charge of an event. This is done because quarks and gluons coming from hard process scattering cannot stay free and produce quark-antiquark pair or radiate gluons. The parton shower software simulates this process until partons have reached low energy of the order of 1 GeV.

After all the partons have showered, they go through the process of hadronization to produce final state hadrons. Hadronization is done using the models that are tuned on experimental data. Most of the models used for MC in standard model physics at CMS are

based on the Lund string model in the PYTHIA framework [47]. Most showers are usually collimated and they appear as jets in the final state, but sometimes a highly energetic gluon can radiate a high energy gluon which will eventually make a new jet, it can be either ISR during the hard process, or final state radiation (FSR) during the parton shower. This increases jets multiplicity of an event, since we have four jets in our final state this step is important for accurate modeling of our background processes.

3.3.3 Tunes

In hadron-hadron collisions there are two additional interactions, beam-beam remnants (BBRs) and multiple-parton interactions (MPIs), they are modeled as underlying event (UE). BBR is what left behind by incoming beam hadron and does not take part in ISR or hard scattering, and these remnants need to be color connected with the rest of the event. The MPIs are the additional multiple interaction that a parton can go through in the same hadron-hadron collision, these interactions can be hard or semi-hard. To model UE properly in standard MC event generator, a set of parameters (referred as “tune”) are needed to be adjusted so that it provides better fit to the experiment data. CMS made use of the tune `TuneCUETP8M1` [48] (2016 MCs samples) and `TuneCP5` [49] (2017 and 2018 MCs samples) in the PYTHIA for modeling UE.

CHAPTER 4

VBS MEASUREMENT IN ZVJJ FINAL STATE

As discussed in Section 1.2 this analysis targets VBS in the ZV channel accompanied with two jets at large η in final state. The goal of the analysis is to reduce contribution from background processes as much as possible without losing much of signal, and measure signal strength and significance.

Since in our chosen topology Z decays leptonically and V is decaying hadronically, the phase-space of this analysis can be either l^+l^-jjjj or l^+l^-Jjj , where l are leptons (electrons or muons), j are narrow jets and J is a boosted (wider) jet. The phase-space is divided into two broad regions, signal and control. The signal region is constructed based on theory such that the signal is maximized, and the control region is orthogonal to the signal region where we expect contributions mostly from background processes. The main purpose of control region in this analysis is to provide normalization factors for the dominant background process in the signal region.

The analysis is performed “blind” to avoid intrinsic bias i.e. until the analysis procedure is finalized, the collision data is only used in control regions. Once the analysis technique is optimized using MC samples and validated against collision data in control regions and approved by CMS Physics Group then the results are “un-blinded” i.e. measurements are done using collision data in signal region.

4.1 Dataset and Simulation

As discussed in Section 2.2.7 only events that pass Level-1 Trigger and HLT paths are saved for further processing, MC simulation also has these identical steps during event generation to mimic Level-1 Trigger and HLT paths.

CMS processes the datasets centrally and provides various tiers of datasets such as “RECO” datasets, which contains reconstructed objects and no skimming. Average size of an event saved at “RECO” tier is 480 kB per event and on average an analysis will process 3 billion events, which makes this tier not practical in terms of computer processing time and storage if each analysis starts from “RECO”. CMS centrally processes these datasets further and removes certain objects or features to reduce the average event size but still retain enough information for the majority of the analyses.

This analysis uses one such reduced dataset: NANOAOD tier with version “v7” of datasets which has average event size of 2–3 kB.

4.1.1 Data

The collision data events used in this analysis are all certified by CMS Data Quality Monitoring (DQM) and Data Certification (DC) group. The primary trigger object in HLT paths are leptons p_T threshold, and since in our final state we are looking for Z boson decaying into two leptons, we require single and double lepton triggers for our analysis. Depending on the detector conditions and LHC storage capacity we have slightly different thresholds in triggers across different years. Table 4.1 contains the list of HLT paths used in this analysis.

Table 4.1: Trigger paths used to select events in CMS collision data

Dataset	Year	HLT Path	p_T	Threshold
Single Muon	2016	HLT_IsoMu24		24 GeV
		HLT_IsoTkMu24		24 GeV
	2017	HLT_IsoMu27		27 GeV
	2018	HLT_IsoMu24		24 GeV
Single Electron	2016	HLT_Ele27_WPTight_Gsf		27 GeV
		HLT_Ele25_eta2p1_WPTight_Gsf		25 GeV
	2017	HLT_Ele35_WPTight_Gsf		35 GeV
	2018	HLT_Ele32_WPTight_Gsf		32 GeV
Double Muon	2016	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ		17, 8 GeV
		HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ		17, 8 GeV
		HLT_TkMu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ		17, 8 GeV
	2017	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8		17, 8 GeV
		HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8		17, 8 GeV
	2018	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8		17, 8 GeV
Double Electron	2016	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ		23, 12 GeV
		HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL		23, 12 GeV
	2017	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL		23, 12 GeV
	2018	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL		23, 12 GeV

4.1.2 MC Simulations

The EW VBS signal is generated with MADGRAPH5+PYTHIA8 at LO with α_{EW}^6 i.e. all vertices in the tree level Feynman diagram are EW vertices. The QCD induced VBS background process which is very similar to our signal is generated with same configuration but with $\alpha_{EW}^4 \alpha_{QCD}^2$ i.e. two of the six vertices are QCD. The dominant backgrounds to the analysis are Drell-Yan (DY) + Jets, top-quark based processes consisting of single top-quark (t and s -channel), single top-quark in association with W boson (tW) and top-quark pair ($t\bar{t}$) production. The DY + Jets are generated at LO with MADGRAPH5+PYTHIA8

in bins of HT (scalar sum of all the jet transverse momentums in the event), to have more statistics for higher HT bins. Top-quark background process are generated at next-to-next-to-leading order (NNLO), s -channel is generated with `MADGRAPH5_aMC@NLO +PYTHIA8` and others: t -channel, tW , $t\bar{t}$ are generated with `POWHEG+PYTHIA8` [50]. The QCD multijet background is small and not considered in this analysis, since we are constructing Z candidate from two leptons and requiring mass to be within [76,106] GeV mass window (Section 4.2). Tables 4.2, 4.3, 4.4 contains the complete list of Signal and Background MC samples used for modeling in this analysis.

Table 4.2: List of MC samples for Signal and Background modeling

Process	Year	Dataset Name	Cross Section (pb)	Events Generated
VBS_EWK (Signal)	2016	WminusTo2JZTo2LJJ_dipoleRecoil_EMK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.02982	509444
	2016	WplusTo2JZTo2LJJ_dipoleRecoil_EMK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.05401	579648
	2016	ZTo2LZTo2JJJ_dipoleRecoil_EMK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.01589	288106
	2017, 2018	WminusTo2JZTo2LJJ_dipoleRecoil_EMK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.02982	586811, 592587
	2017, 2018	WplusTo2JZTo2LJJ_dipoleRecoil_EMK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.05401	594517, 598523
	2017, 2018	ZTo2LZTo2JJJ_dipoleRecoil_EMK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.01589	287660, 297577
VBS_QCD (Background)	2016	WminusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8	0.3488	500000
	2016	WplusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8	0.575	500000
	2016	ZTo2LZTo2JJJ_QCD_LO_SM_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8	0.3449	100000
	2017, 2018	WminusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.3488	1476120, 1479242
	2017, 2018	WplusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.575	1428816, 1451346
	2017, 2018	ZTo2LZTo2JJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.3449	149536, 143152
DY + Jets LO (Background)	2016	DYJetsToLL_M-50_HT-70to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	169.9	9691660
	2016	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	147.4	11017086
	2016	DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	41.04	9009137
	2016	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.674	9725661
	2016	DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.358	8292957
	2016	DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.6229	2673066
	2016	DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.1512	596079
	2016	DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.003659	399492
	2017	DYJetsToLL_M-50_HT-70to100_TuneCP5_13TeV-madgraphMLM-pythia8	167.33	933543
	2017	DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	161.1	15124171
	2017	DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	48.66	11896758
	2017	DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	6.968	11294006
	2017	DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	1.743	8691608
	2017	DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	0.8052	3089712
	2017	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.1933	616923
	2017	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.003468	401334

Table 4.3: List of MC samples for Signal and Background modeling

Process	Year	Dataset Name	Cross Section (pb)	Events Generated
DY + Jets LO (Background)	2018	DYJetsToll_M_50_HT_70to100_TuneCP5_PSweights_13TeV-madgraphMLM_pythia8	167.33	10010341
	2018	DYJetsToll_M_50_HT_100to200_TuneCP5_PSweights_13TeV-madgraphMLM_pythia8	161.1	11516746
	2018	DYJetsToll_M_50_HT_200to400_TuneCP5_PSweights_13TeV-madgraphMLM_pythia8	48.66	10840079
	2018	DYJetsToll_M_50_HT_400to600_TuneCP5_PSweights_13TeV-madgraphMLM_pythia8	6.968	18902490
	2018	DYJetsToll_M_50_HT_600to800_TuneCP5_PSweights_13TeV-madgraphMLM_pythia8	1.743	8826238
	2018	DYJetsToll_M_50_HT_800to1200_TuneCP5_PSweights_13TeV-madgraphMLM_pythia8	0.8052	3120982
	2018	DYJetsToll_M_50_HT_1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM_pythia8	0.1933	531566
	2018	DYJetsToll_M_50_HT_2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM_pythia8	0.003468	415517
	2016	ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV_PSweights_powheg_pythia8	38.06	2710849
	2016	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	34.97	174109584
Top (Background)	2016	ST_tW_top_5f_NoFullyHadronicDecays_13TeV_PSweights_powheg-pythia8	38.09	3213335
	2016	ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	34.91	173908704
	2016	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	67.91	17771480
	2016	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	11.33	31835782
	2016	ST_s-channel_4f_leptonDecays_13TeV_PSweights_amcatnlo-pythia8	3.365	33572360
	2016	ST_s-channel_4f_hadronicDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	11.24	54761756
	2016	ST_s-channel_4f_InclusiveDecays_13TeV-amcatnlo-pythia8	10.12	29561764
	2016	TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8	377.96	21432309760
	2016	TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8	365.34	32266942208
	2016	TTTo2l2Nu_TuneCP5_PSweights_13TeV_powheg_pythia8	88.29	4891619840
TTToSemiLeptonic (Background)	2017	TTToSemiLeptonic_TuneCP5_13TeV-powheg_pythia8	365.34	72526299136
	2017	TTTo2l2Nu_TuneCP5_13TeV-powheg_pythia8	86.99	648729856
	2017	TTToHadronic_TuneCP5_13TeV-powheg_pythia8	377.96	61926825984
	2017	ST_s-channel_antitop_leptonDecays_13TeV_PSweights_powheg_pythia	1.33	3422897
	2017	ST_s-channel_top_leptonDecays_13TeV_PSweights_powheg_pythia	2.13	12862777
	2017	ST_t-channel_antitop_5f_TuneCP5_PSweights_13TeV-powheg-pythia8	27.19	557503744
	2017	ST_t-channel_top_5f_TuneCP5_13TeV-powheg_pythia8	45.7	1418556928
	2017	ST_tW_antitop_5f_InclusiveDecays_TuneCP5_13TeV-powheg_pythia8	12.04	279005344
	2017	ST_tW_top_5f_InclusiveDecays_TuneCP5_13TeV-powheg_pythia8	12.04	272081088

Table 4.4: List of MC samples for Signal and Background modeling

Process	Year	Dataset Name	Cross Section (pb)
Top (Background)	2018	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	365.34
	2018	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	86.99
	2018	TTToHadronic_TuneCP5_13TeV-powheg-pythia8	377.96
	2018	ST_s-channel_antitop_leptonDecays_13TeV-PSweights_powheg-pythia	1.33
	2018	ST_s-channel_top_leptonDecays_13TeV-PSweights_powheg-pythia	2.13
	2018	ST_t-channel_antitop_5f_TuneCP5_13TeV_powheg-pythia8	27.19
	2018	ST_t-channel_top_5f_TuneCP5_13TeV_powheg-pythia8	45.7
	2018	ST_tW_DS_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	12.04
	2018	ST_tW_DS_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	12.04

4.2 Event Selection

In the first stage of selection, events are selected if there are a minimum number of objects for the relevant analysis categories i.e. at least two leptons of same flavor ($p_T > 10 \text{ GeV}$), and either four narrow jets (AK4) (Resolved ZV category) or two narrow jets (AK4) plus one wider jet (AK8) (Boosted ZV category).

After this initial skimming step, tight leptons (muons and electrons) are selected. If there are exactly two tight leptons of same flavor, and opposite charge then the event is selected, and a Z candidate is constructed from these two OSSF leptons. Events with Z candidate mass in the range [76, 106] GeV are selected.

Following are the tight selections for the muons and electrons:

- **Tight Muons:** Muons with $p_T > 20 \text{ GeV}$, passing tight ID [35], “pfRelIso04” (Section 3.2.1) less than 0.15 and impact parameters $d_{xy} < 0.01, d_z < 0.1$ [35].
- **Tight Electrons:** Electrons with $p_T > 20 \text{ GeV}$ and MVA based ID depending on year of the dataset. For 2016 dataset: passing “mvaSpring16GP_WP90” (MVA based ID, 90% efficiency) and “pfRelIso03_all” less than 0.0571 for $|\eta| > 1.479$ and less than 0.0588 otherwise. For 2017 and 2018 dataset: passing “mvaFall17V2Iso_WP90” (MVA based ID + Isolation, 90% efficiency) and “pfRelIso03_all” less than 0.06 is required [37].

4.2.1 VBS Tagged Jets and V Jet Candidate

After selecting leptons and constructing the Z candidate, we look for the vector boson V as a “FatJet” (merged jet, AK8) candidate with $p_T > 200 \text{ GeV}$, within $|\eta| < 2.4$ and in softdrop (Section 3.2.3.2) mass range of [40, 150] GeV. Then all the FatJets are checked

for isolation from selected leptons with $\Delta R = 0.8$ and then N-subjetiness (Section 3.2.3.1) $\tau_{21} < 0.45$ cut is applied. Finally the FatJet with mass nearest to the average mass of W and Z is selected.

- **Boosted:** If after the above described selection, there is a FatJet candidate passing all the cuts for V then the event is categorized as “boosted”.
- **VBS Jets:** Whether the event is a boosted event or not, narrow jets (AK4) are considered as VBS jets. This is done by first cleaning jets by applying tight “JetID”, tight “pileup JetID” (PUID) [41], $p_T > 30$, and checking isolation against selected leptons with $\Delta R = 0.4$, selected FatJet (if any) with $\Delta R = 0.8$ and other jets with $\Delta R = 0.4$. Then all the combinations of jets are checked as pairs, to find the pair with highest invariant mass with minimum of 500 GeV and pseudorapidity difference ($|\Delta\eta|$) more than 2.5. The jets forming that pair are tagged as VBS jets. If no such pair is found, the event is rejected.
- **Resolved:** If there is no boosted event but there are VBS tagged jets, then from the rest of jet collection, a pair is searched with each jet within $|\eta| < 2.4$, and invariant mass nearest to the average mass of W and Z and in the range of [40, 150] GeV. Jets from the selected pair are then labelled as V jet candidates and event is categorized as “resolved”.

4.3 Control and Signal Region

To model our background as accurately as possible, the selected events are divided into two regions based usually on an observed property of the selected objects.

For this analysis, the mass window of V is selected to define the region boundaries and two regions are defined as:

- **Signal Region:** $65 < M_V < 105$ GeV, it's an 20 GeV window around the mass of V .

This is a narrow mass window, and we expect higher signal sensitivity in this region.

- **Control Region:** $40 < M_V < 65$ and $105 < M_V < 150$ GeV, these mass sidebands around signal region are rich in background. This region is dominated by DY plus jets events and will be used to model and normalize this background process.

4.3.1 MC Scale Factors

Whenever we apply some cut on object properties on data and MC, the efficiency of the cut can be different for the data than it is for MC. To account for this difference in efficiency of data and MC, scale factors are calculated and applied to MC. The scale factors are usually function of p_T and η of the object being used for selection. These scale factors are then applied as event weight i.e. they are multiplied to the cross-section weight factor applied to the MC events. Following are the scale factors used in this analysis:

- **Pile-Up Reweighting:** During run2 data taking at CMS the average number of pileup events in a collision can vary from 5 to 100. To have the same shape for this variable in MC, a reweighting scale derived from the data-MC difference is applied to MC.
- **Trigger Efficiency:** Since the HLT trigger paths are defined by the p_T threshold of the object in the trigger, they can have different efficiency in data and MC. Scale factors are applied to MC as a function of p_T and η to correct for data-MC difference.

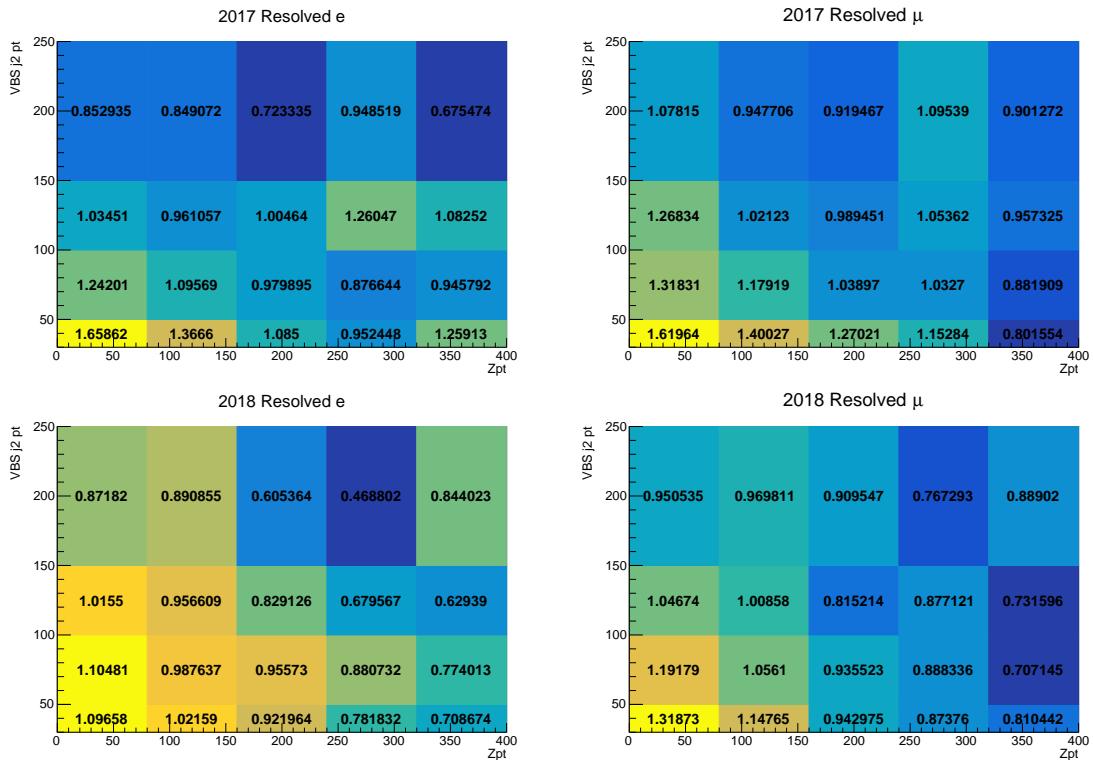
- **Lepton Identification/Isolation:** Scale factors to account data-MC difference arising from tight selection of electrons and muons are derived from “tag and probe” method and applied to the MC.
- **Tight Pile-Up Jet Identification:** To reduce jets coming from pileup vertices, tight Pile-Up Identification is applied to the VBS tagged jets. Scale factors are calculated for both VBS jets and applied to MC.
- **L1 Prefire Correction:** During 2016–2017 operations ECAL $|\eta| > 2.5$ region wrongly associated energy deposit from previous bunch crossing causing L1 Trigger system to “prefire” and accept the event. L1 Trigger prefiring scale factors recommended by CMS for 2016 and 2017 year datasets [51].

4.3.2 DY+Jets Normalization

DY+Jets process MC is generated at LO. The MC doesn’t model well the rather peculiar subset of the DY+Jets that is relevant for this analysis. To normalize and model our main background MC DY+Jets, we extract scale factors from bins of variables in control region and apply it to both control and signal region. For boosted category of all years 2016, 2017, 2018 and resolved in 2016 1D bins of p_T of Z were used. For resolved category in 2016 and 2017 2D bins of p_T of Z and VBS trailing jet were used. Table 4.5 lists the normalization factors extracted from the 1D bins of p_T of Z . Figure 4.1 shows the normalization factors extracted from the 2D bins of p_T of Z and VBS trailing jet.

Table 4.5: Normalization factors from 1D bin of p_T of Z

Z p_T bin	Resolved				Boosted			
	2016		2016		2017		2018	
	e	μ	e	μ	e	μ	e	μ
[0, 80]	1.11	1.19	1.09	1.03	0.92	1.02	0.86	1.01
[80, 160]	1.07	1.12	1.00	1.03	0.83	0.84	0.86	0.81
[160, 240]	1.14	1.13	1.04	0.88	0.82	0.80	0.75	0.68
[240, 320]	0.93	1.14	0.83	1.15	0.99	0.93	0.73	0.76
[320, 400]	1.07	1.08	0.73	0.82	1.23	0.85	0.70	0.76
[400, 480]	1.02	1.06	0.62	0.70	0.91	0.84	0.67	0.67
[480, inf]	0.84	0.81	0.40	0.56	0.45	0.76	0.64	0.68

Figure 4.1: Normalization factors from 1D bin of p_T of Z and VBS trailing jet. Top to Bottom: 2017 and 2018 resolved. Left to Right: electron and muon channel.

4.4 Kinematics Distributions

To understand how the kinematics of various final state objects and the candidate constructed from them behave, they are studied in the control region by stacking up all the background histograms and overlaying them with the data as points. Signal MC is also overlayed on the histograms, since it's small, for visualization purposes it's scaled with a factor shown in the legends of the plots.

The kinematics plots also include a ratio plot of data and total background MC, and uncertainties in histograms of MC are shown as shaded region which includes total statistical bin error, total JES and theoretical renormalization (QCD scale) error on the DY+Jets and VBS_QCD MC samples.

Section 4.4.1 and 4.4.2 shows kinematics plots in control region for boosted and resolved ZV category respectively. These distributions show reasonable agreement of data-to-MC within statistical fluctuations. This shows that our background is well modeled in control region.

4.4.1 Boosted ZV DY+Jets Control Region

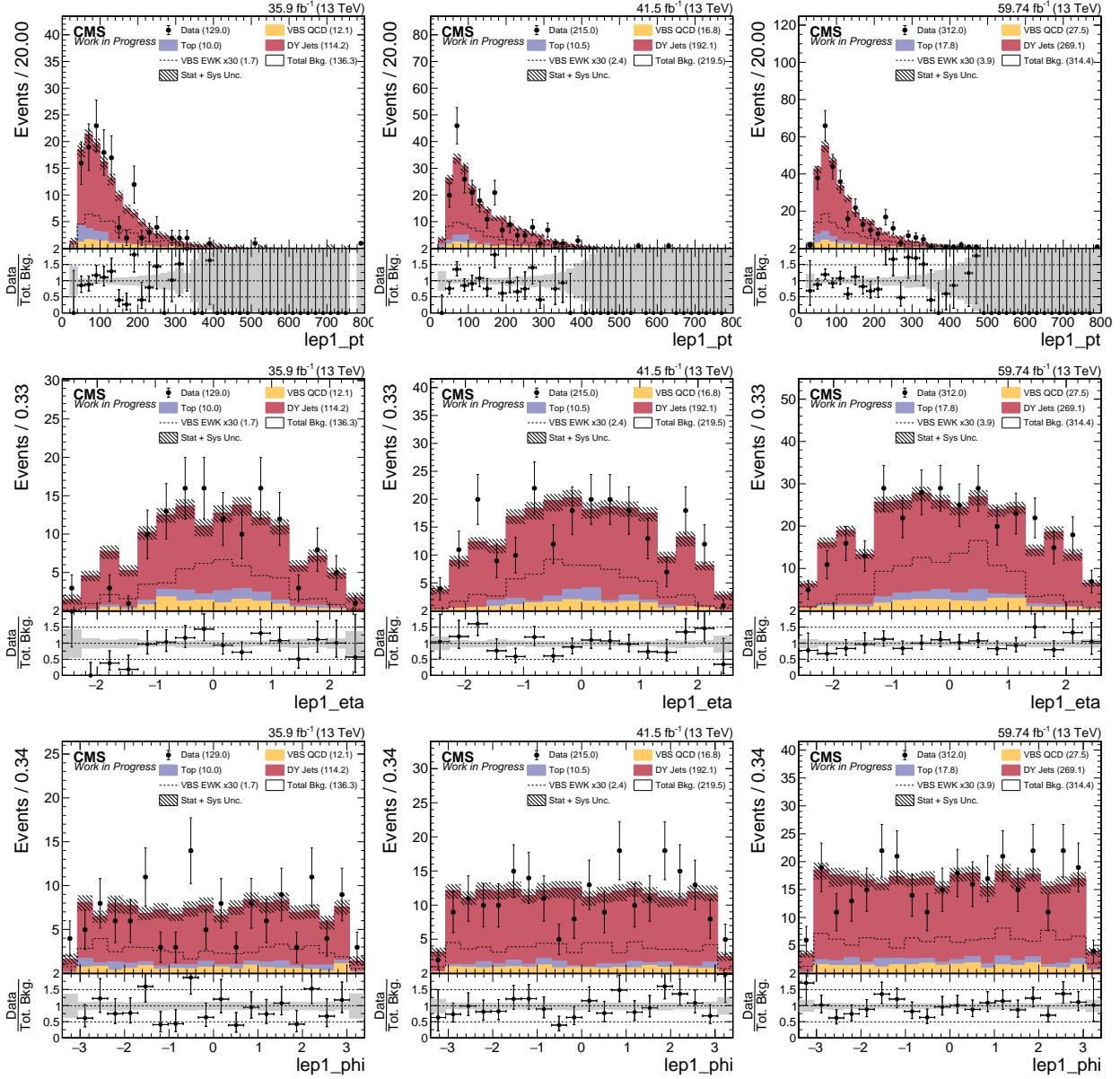


Figure 4.2: DY+Jets Control Region: Leading electron kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

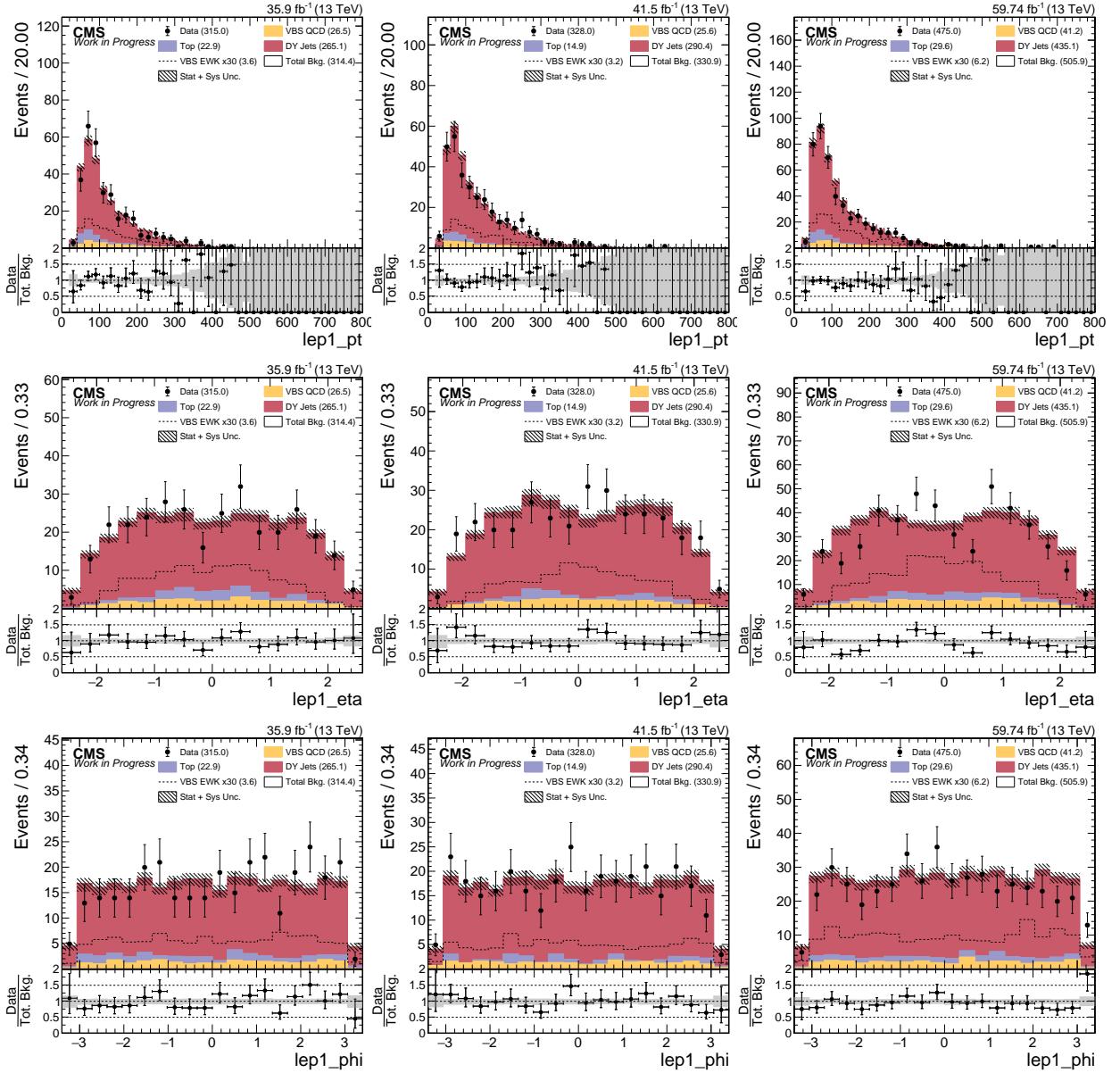


Figure 4.3: DY+Jets Control Region: Leading muon kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

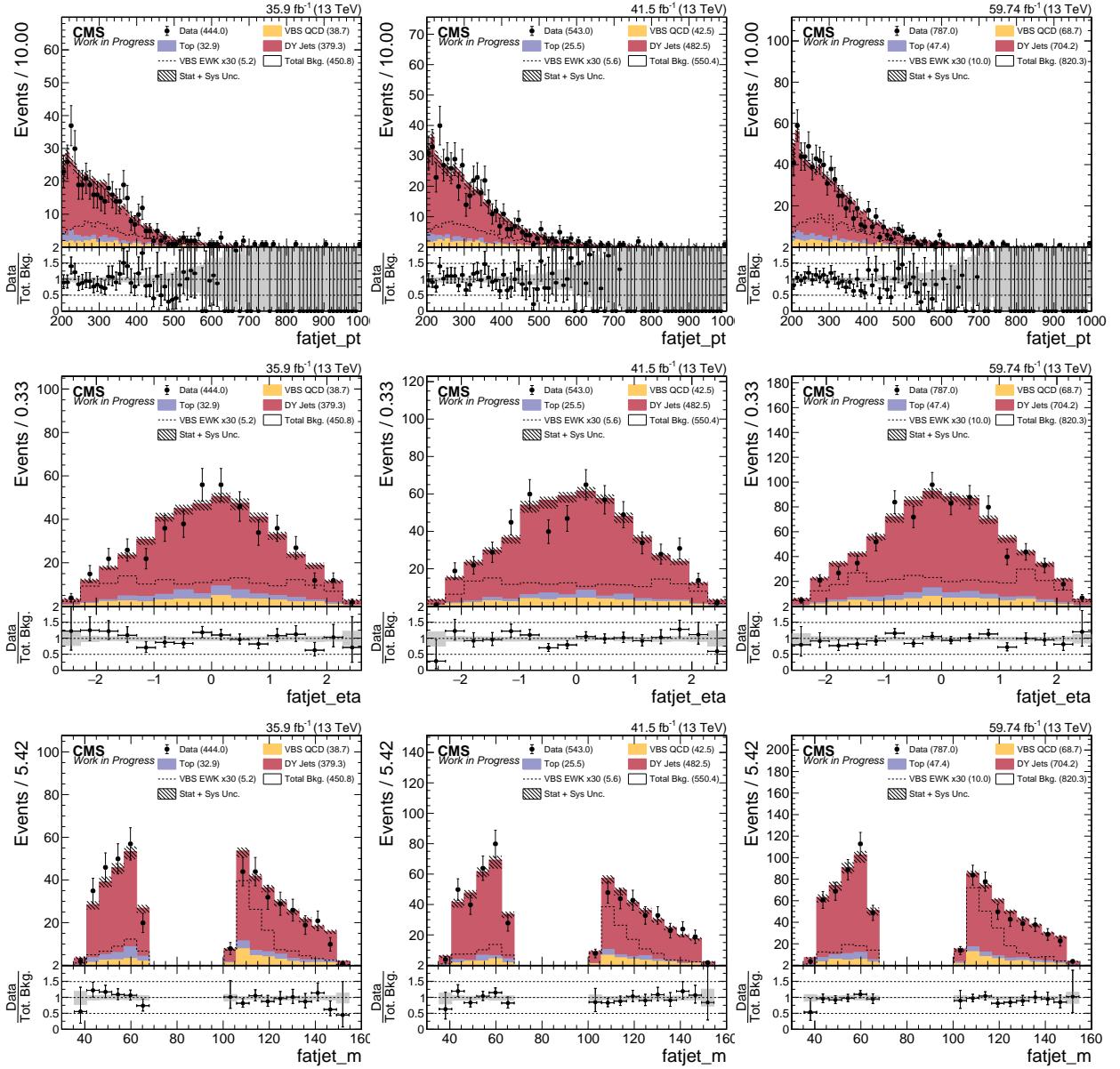


Figure 4.4: DY+Jets Control Region: Hadronic boson kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and mass m .

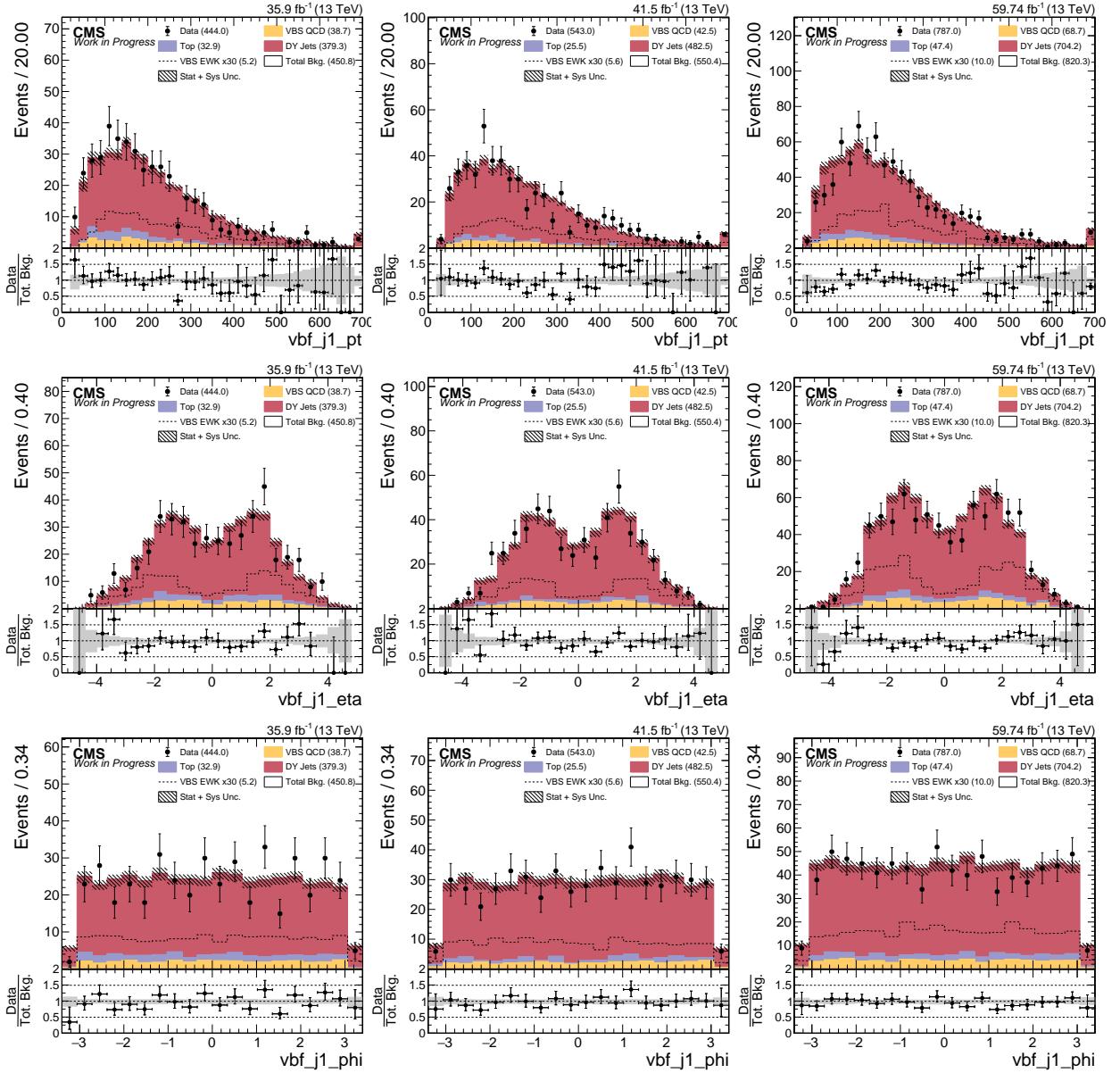


Figure 4.5: DY+Jets Control Region: Leading VBS tagged jet kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

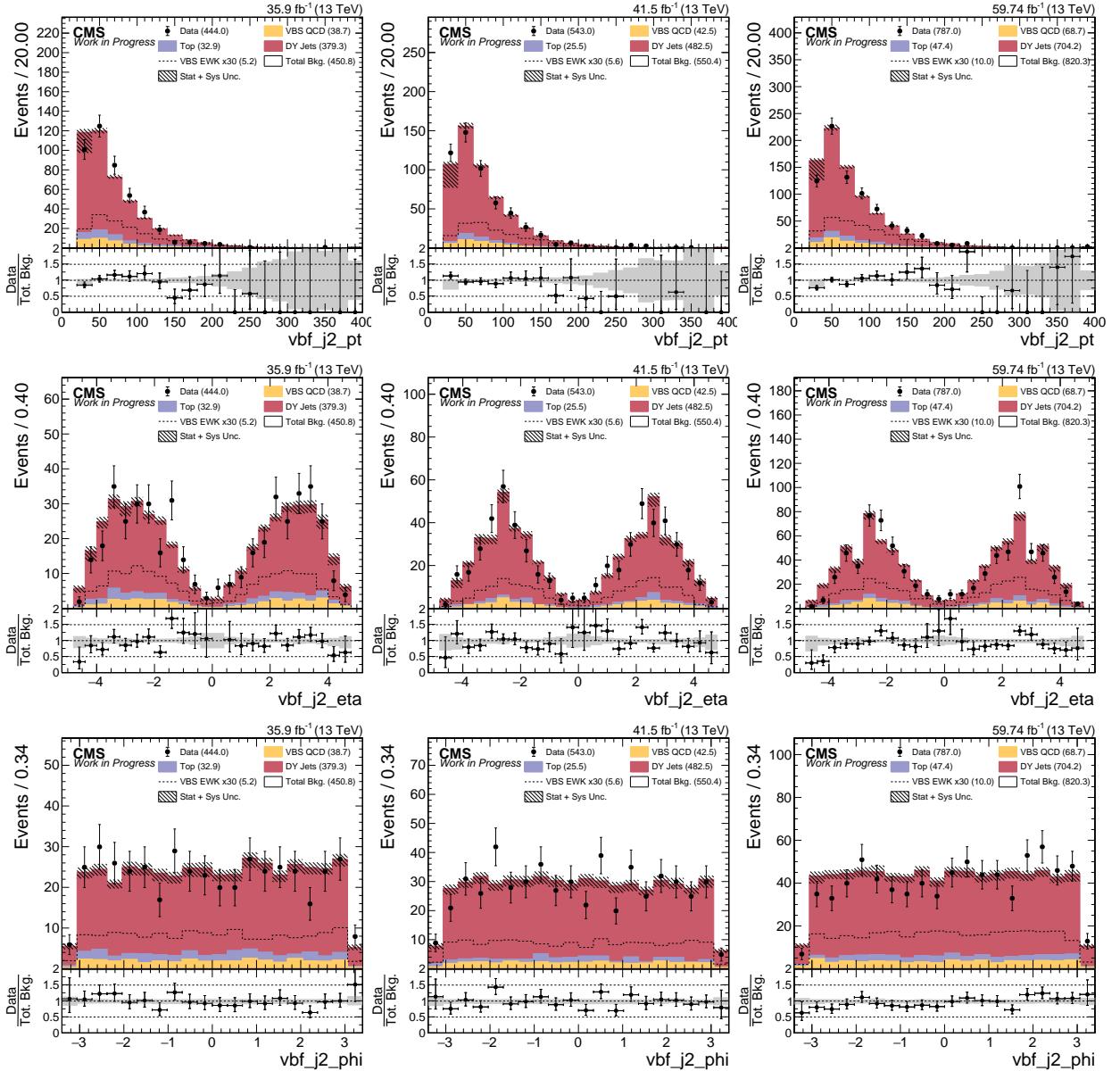


Figure 4.6: DY+Jets Control Region: Trailing VBS tagged jet kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

4.4.2 Resolved ZV DY+Jets Control Region

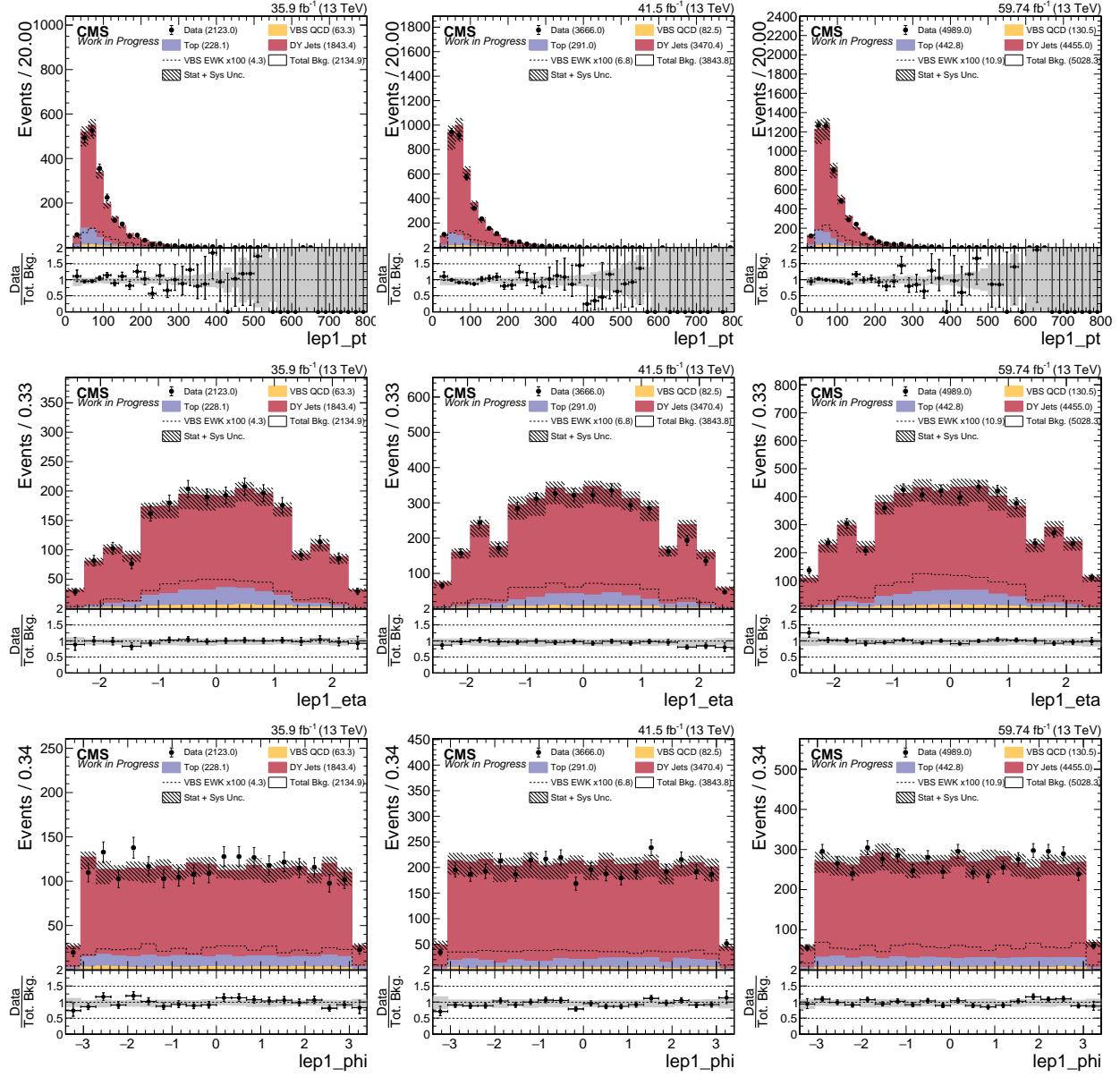


Figure 4.7: DY+Jets Control Region: Leading electron kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

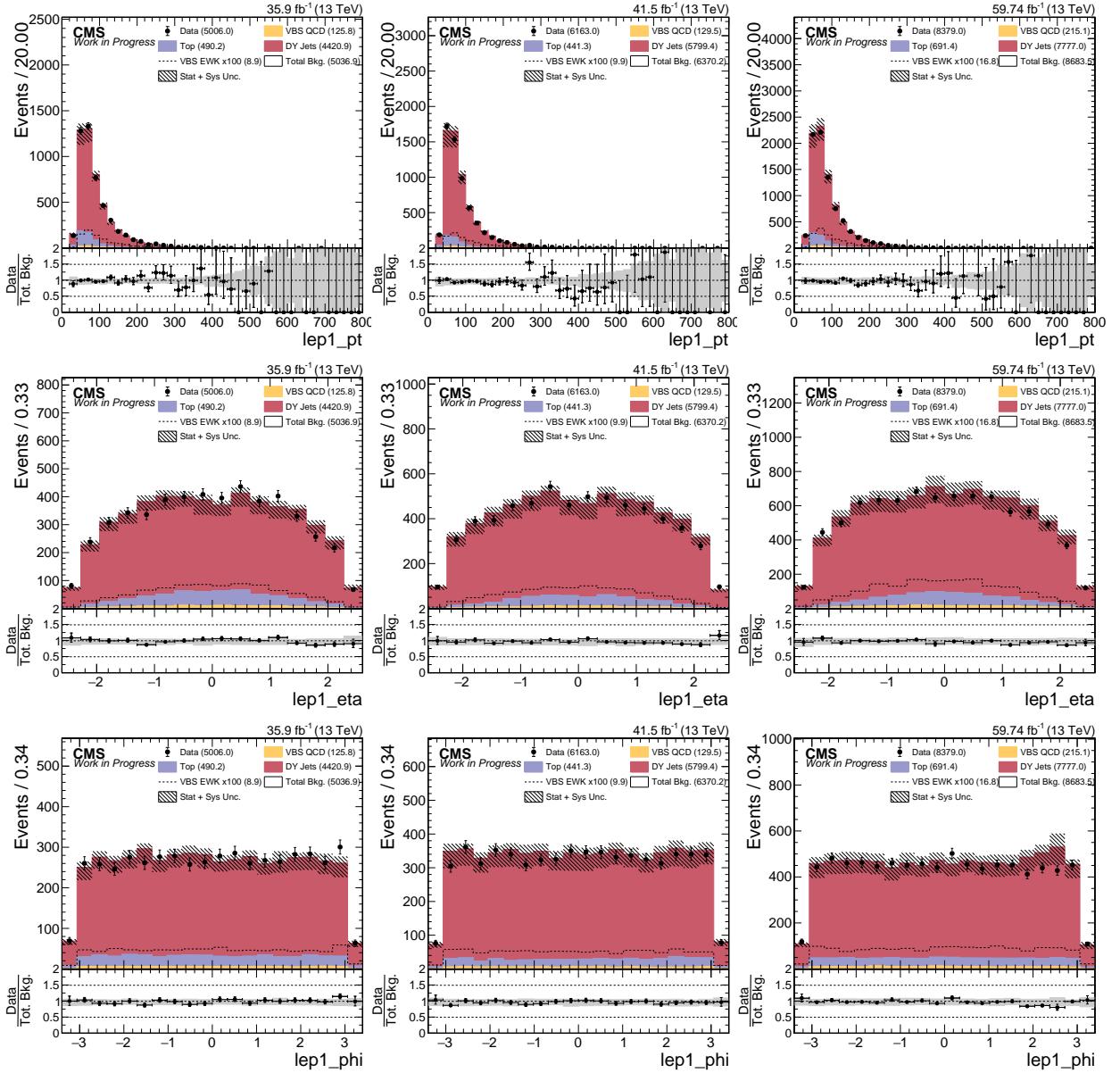


Figure 4.8: DY+Jets Control Region: Leading muon kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

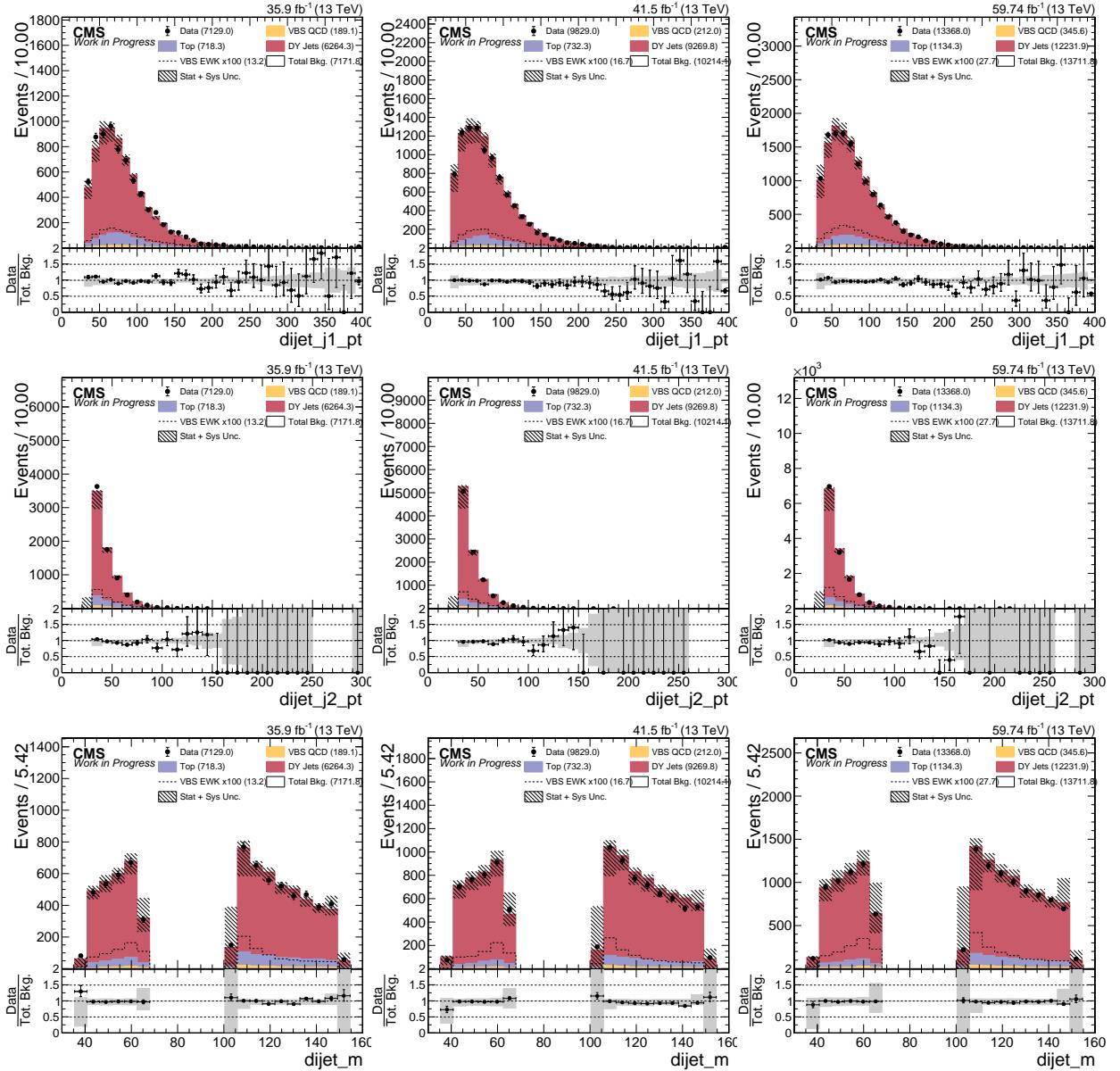


Figure 4.9: DY+Jets Control Region: Hadronic boson leading and trailing jet p_T in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T of leading jet, p_T of trailing jet, and invariant mass m_{jj} .

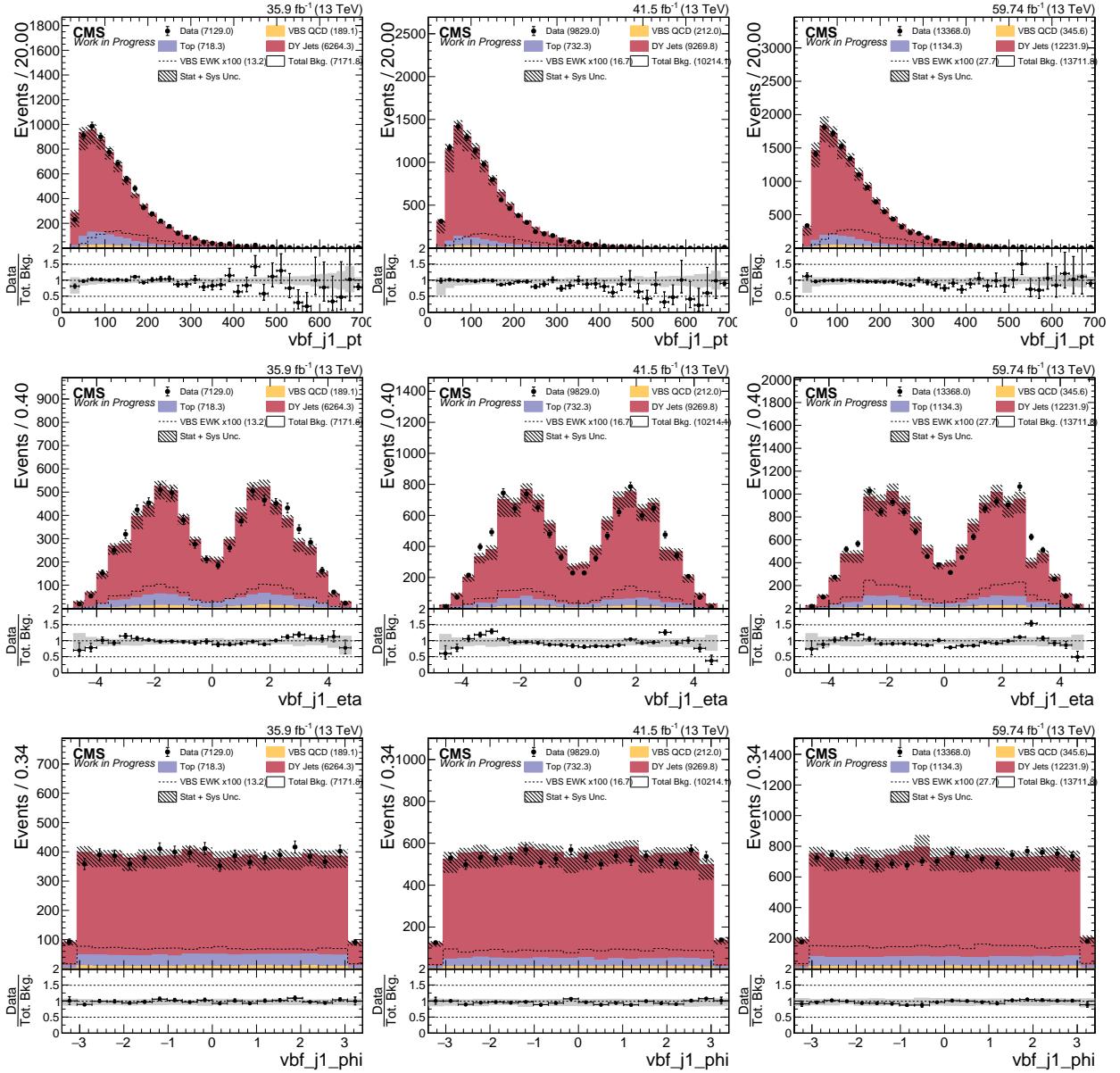


Figure 4.10: DY+Jets Control Region: Leading VBS tagged jet kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

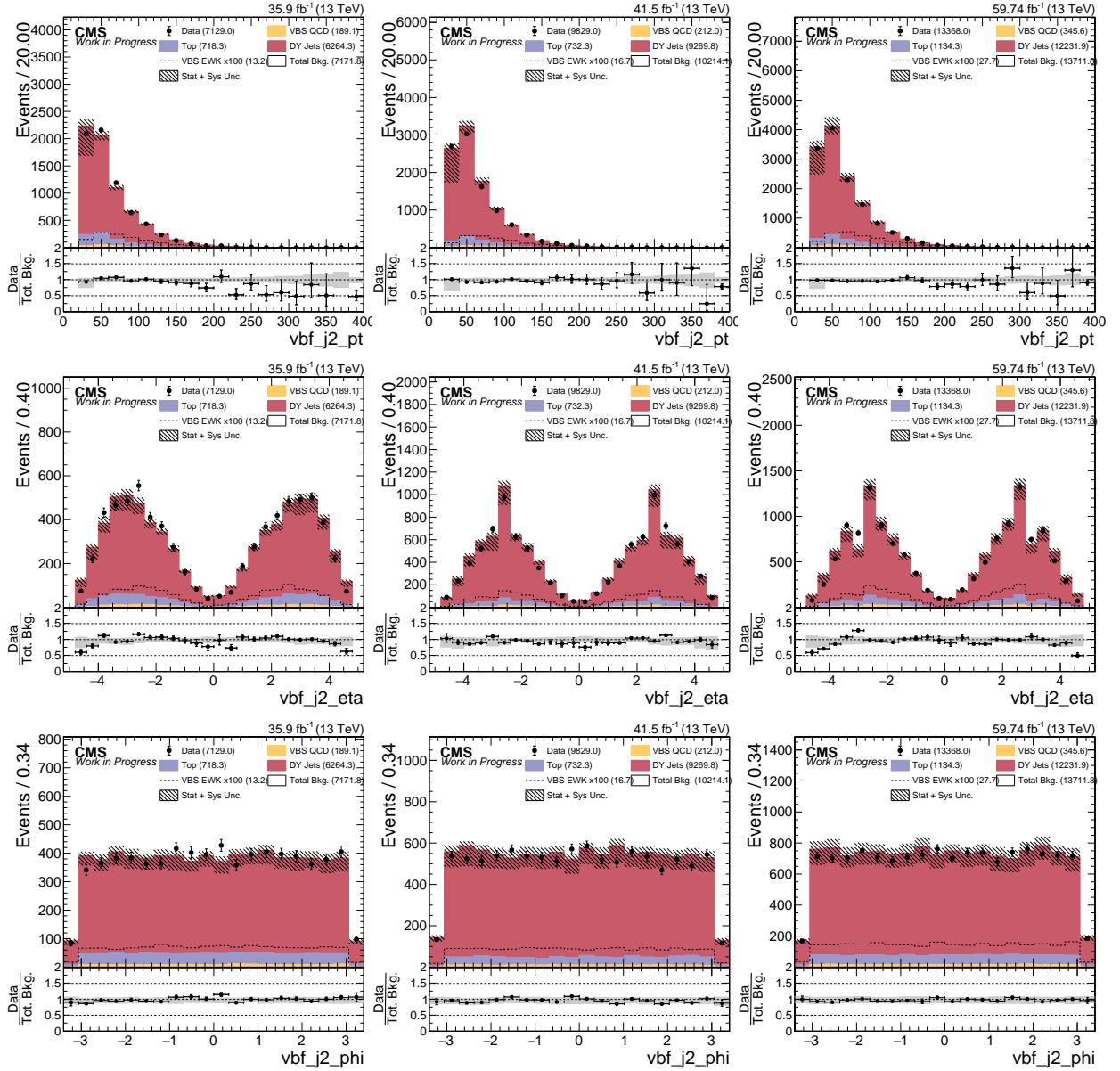


Figure 4.11: DY+Jets Control Region: Trailing VBS tagged jet kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

4.5 Machine Learning Modeling

Instead of traditional cut-based analysis, we decided to use multivariate analysis (MVA) a.k.a. machine learning (ML) technique to build a signal vs background classifier. The main reasoning behind using a MVA technique is so that we can build a model which can learn our analysis topology from looser selection regions and still let us keep higher statistics for final measurement.

4.5.1 Algorithm: Gradient Boosted Decision Tree

Boosting is a method of combining many weak learners sequentially into single classifier. In boosted decision tree (BDT) boosting is applied to *decision tree*. A decision tree is a binary tree, example shown in Figure 4.12, at each node a decision is made using the input variables which provides the best separation¹ at that node, this process is repeated until events at a node reaches below certain threshold, and then depending on the majority of the events in the final node, the node is classified as signal or background.

There are several boosting algorithms, the one discussed and used in this analysis is the gradient boost. Gradient boost is an additive model of adding decision trees in series and with each tree in the sequence, training on the adjusted dataset to minimize the *loss function* with respect to the previous model. The role of the loss function is to estimate the difference between actual and predicted values. This improves the model sequentially and is faster since only a subset of training dataset is used in subsequent decision trees. Toolkit for Multivariate

¹defined as, $\frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy$, where \hat{y}_S and \hat{y}_B are probability density functions of signal and background.

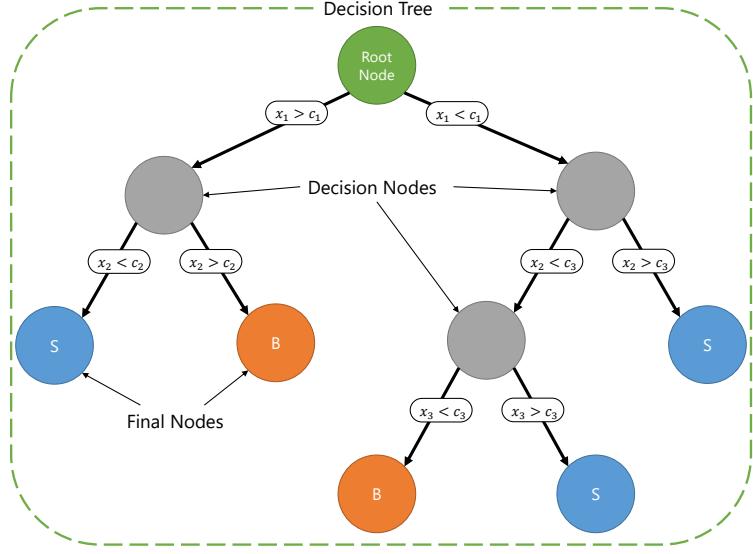


Figure 4.12: Decision Tree.

Analysis (TMVA), part of ROOT package is used in this analysis for implementing the gradient BDT of binomial log-likelihood loss function for classification [52].

4.5.2 Training and Results

Two models were trained for boosted and resolved topology. The training was done using combined MC from all years 2016, 2017, 2018 to benefit from larger statistics (see Table 4.6). Signal MC “VBS_EWK” was trained against background “DY + Jets LO” since that is dominant background in our analysis. The learning process of MVA algorithm is tuned by hyper-parameters values. Each models BDT hyper-parameters were tuned to prevent under and over-fitting. This is done by comparing the signal efficiency for training and testing datasets at three different background rates (1%, 10% and 30%). If the difference is not more than 5%, then the hyper-parameters are acceptable.

Receiver Operating Characteristic (ROC) is a way to visualize performance of signal vs. background efficiency. Area under the curve (AUC) of ROC is a measure of the classifier to

distinguish signal from background. AUC of above 50% means that classifier will be able to distinguish from signal from background events.

Input variables used in each training model were pruned i.e. input variables which do not have significant impact on the performance of the model in terms of AUC of ROC were dropped from the final training models. Distribution of final input variables used in training are shown in Figures 4.13 – 4.18.

Table 4.6: Combined dataset Training and Testing Statistics in Signal Region

Channel	Dataset	Number of Events		
		Training	Testing	Total
Boosted	Signal	7404	7405	14809
Boosted	Background	46991	46991	93982
Resolved	Signal	23425	23425	46850
Resolved	Background	209368	209368	418736

Final input variables used in Boosted and Resolved ZV training model are:

- M_{JJ}^{VBS} : invariant mass of VBS tagged jets.
- Zeppenfeld z^* : Zeppenfeld variable for Z boson [53], $z^* = \frac{\eta_Z - (\eta_{j1}^{VBS} + \eta_{j2}^{VBS})/2}{|\Delta\eta_{jj}^{VBS}|}$
- QGL j_1^{VBS}, j_2^{VBS} : Quark Gluon Likelihood (QGL) of VBS tagged jets. Likelihood whether the jet originated from a quark or gluon.
- M_{VV} invariant mass of Z and V boson.
- HT* (Resolved category only): scalar sum of p_T of jet candidates of V boson and VBS tagged jets.
- η_{lep2} (Resolved category only): η of trailing lepton.

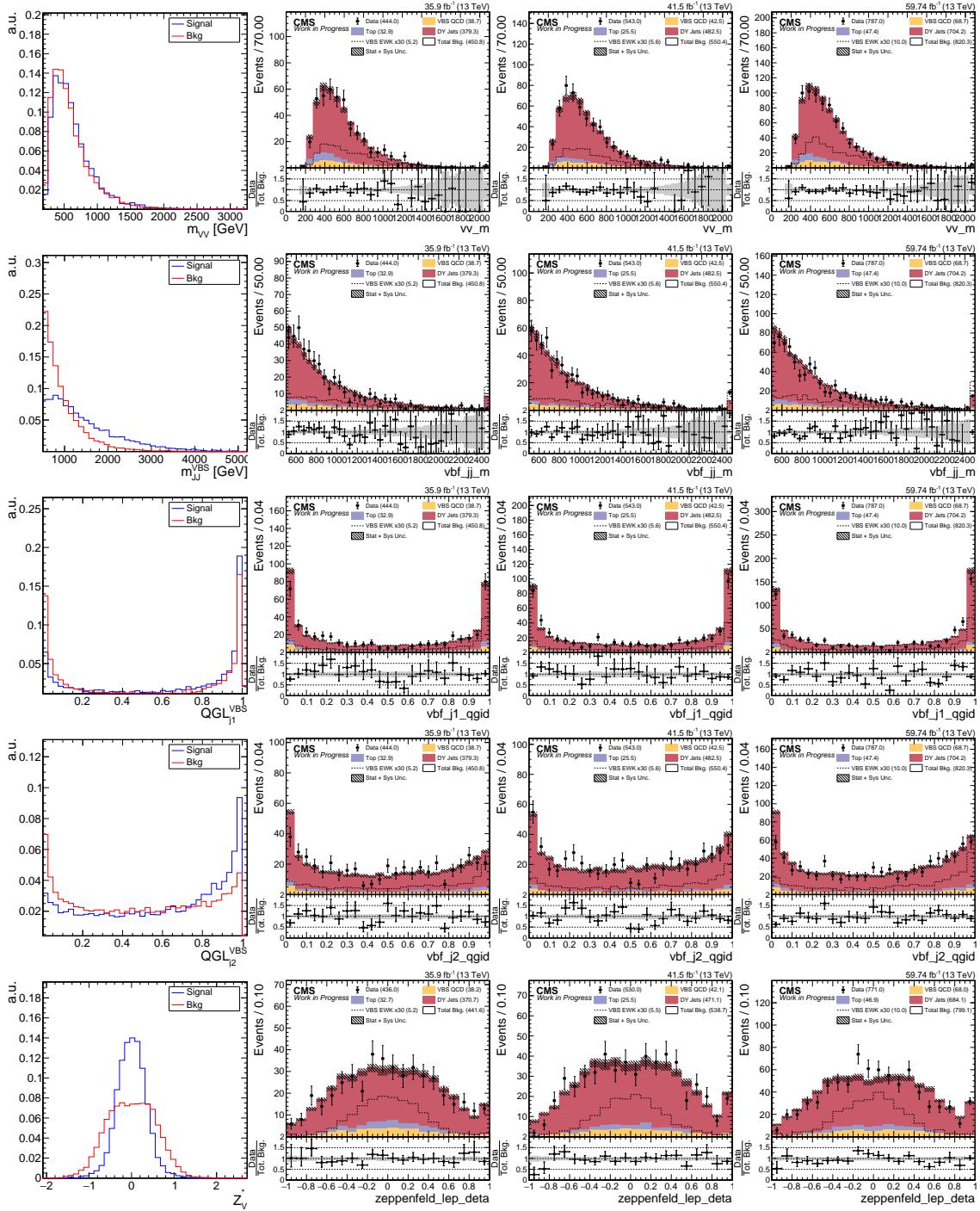


Figure 4.13: Inputs Variables for training Boosted ZV BDT Classifier. Left to Right: Shape comparison of signal and background (combined dataset), control region kinematic distribution in 2016, 2017 and 2018 dataset. Top to Bottom: Diboson invariant mass, VBS tagged jets invariant mass, QGL of leading VBS tagged jet, QGL of trailing VBS tagged jet, Zeppenfeld variable of leptonic boson.

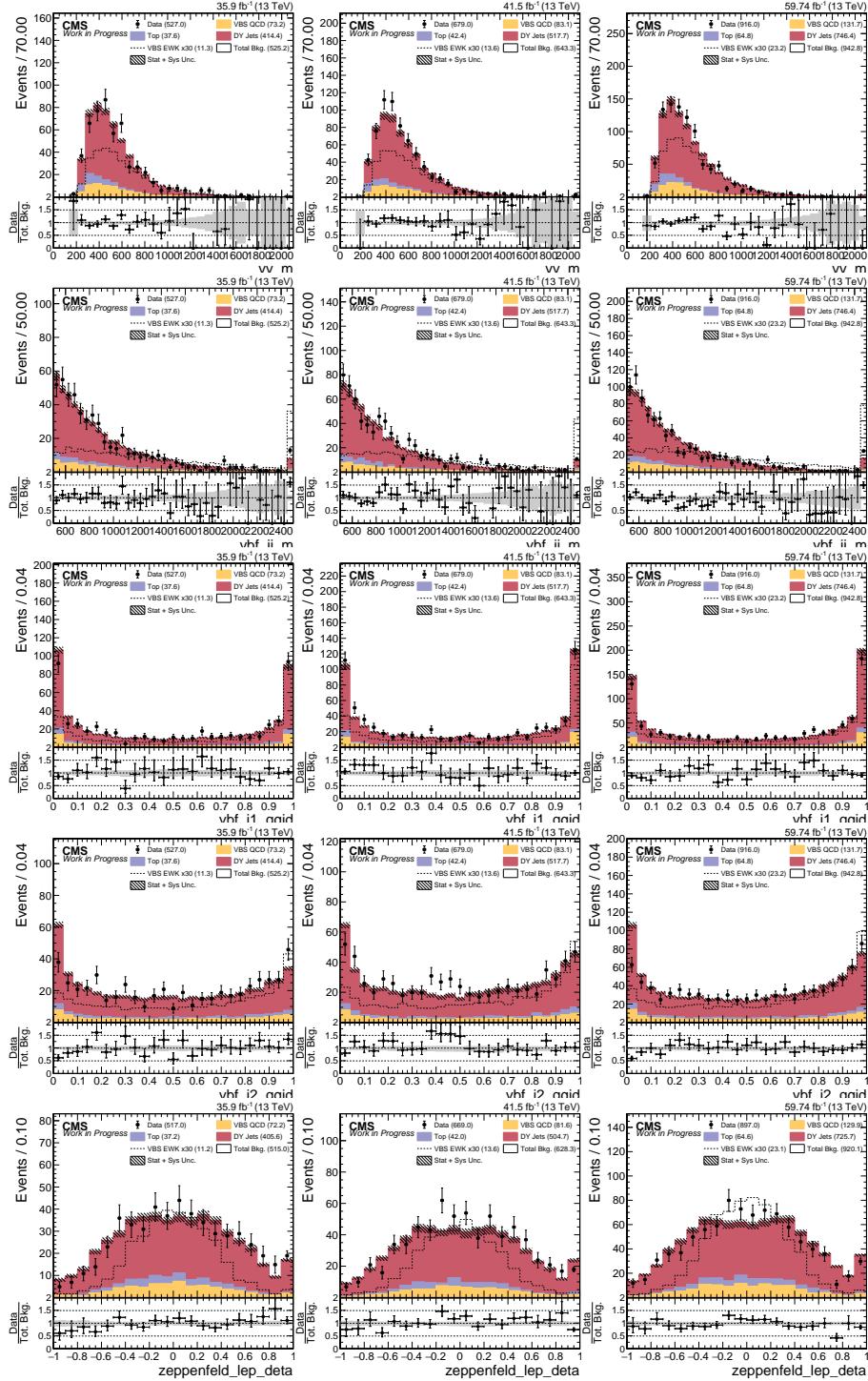


Figure 4.14: Inputs Variables for training Boosted ZV BDT Classifier in signal region. Left to Right: 2016, 2017 and 2018 dataset. Top to Bottom: Diboson invariant mass, VBS tagged jets invariant mass, QGL of leading VBS tagged jet, QGL of trailing VBS tagged jet, Zeppenfeld variable of leptonic boson.

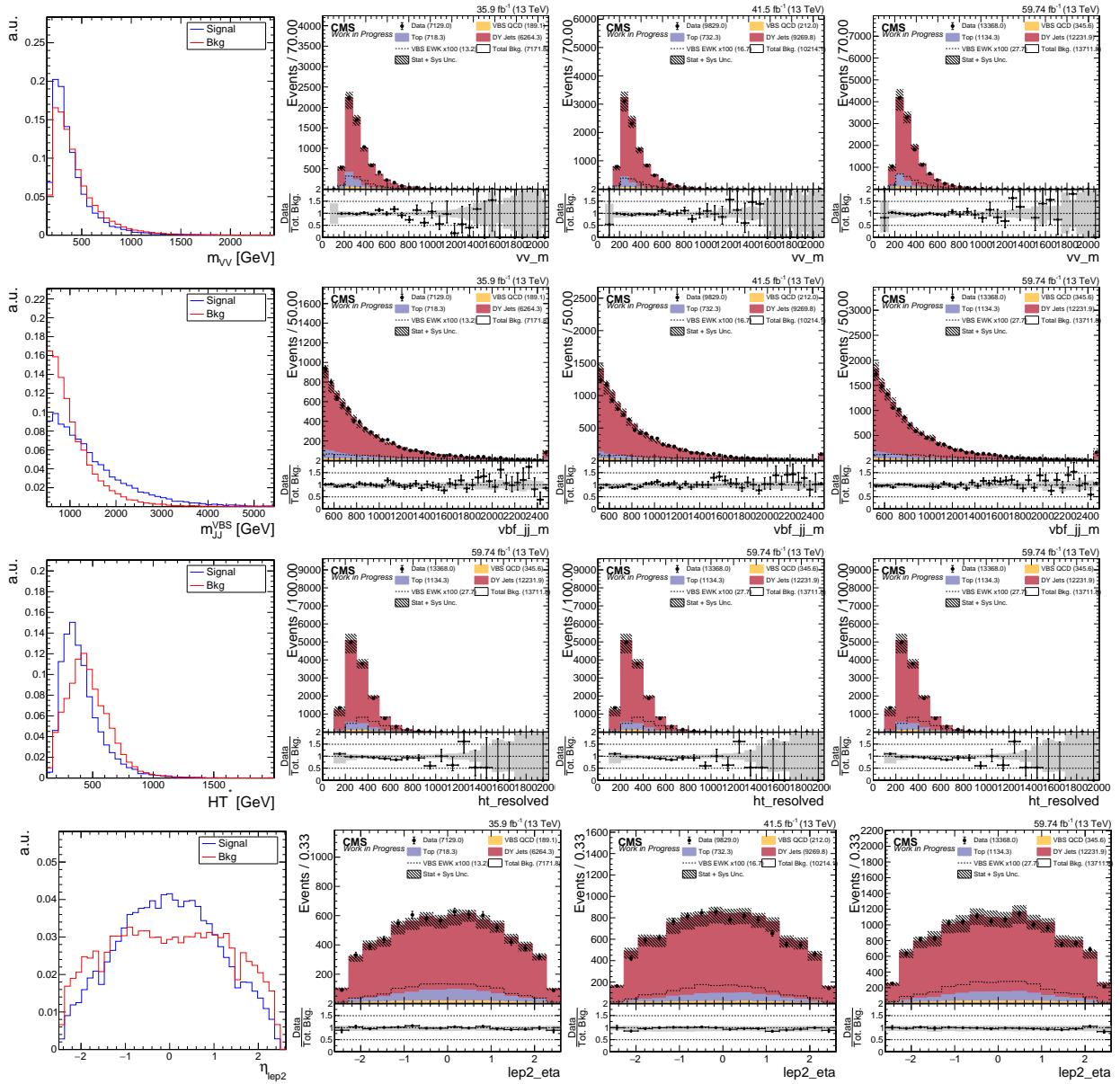


Figure 4.15: Inputs Variables for training Resolved ZV BDT Classifier. Left to Right: Shape comparison of signal and background (combined dataset), control region kinematic distribution in 2016, 2017 and 2018 dataset. Top to Bottom: Diboson invariant mass, VBS tagged jets invariant mass, HT^* (p_T sum of jets), trailing lepton η .

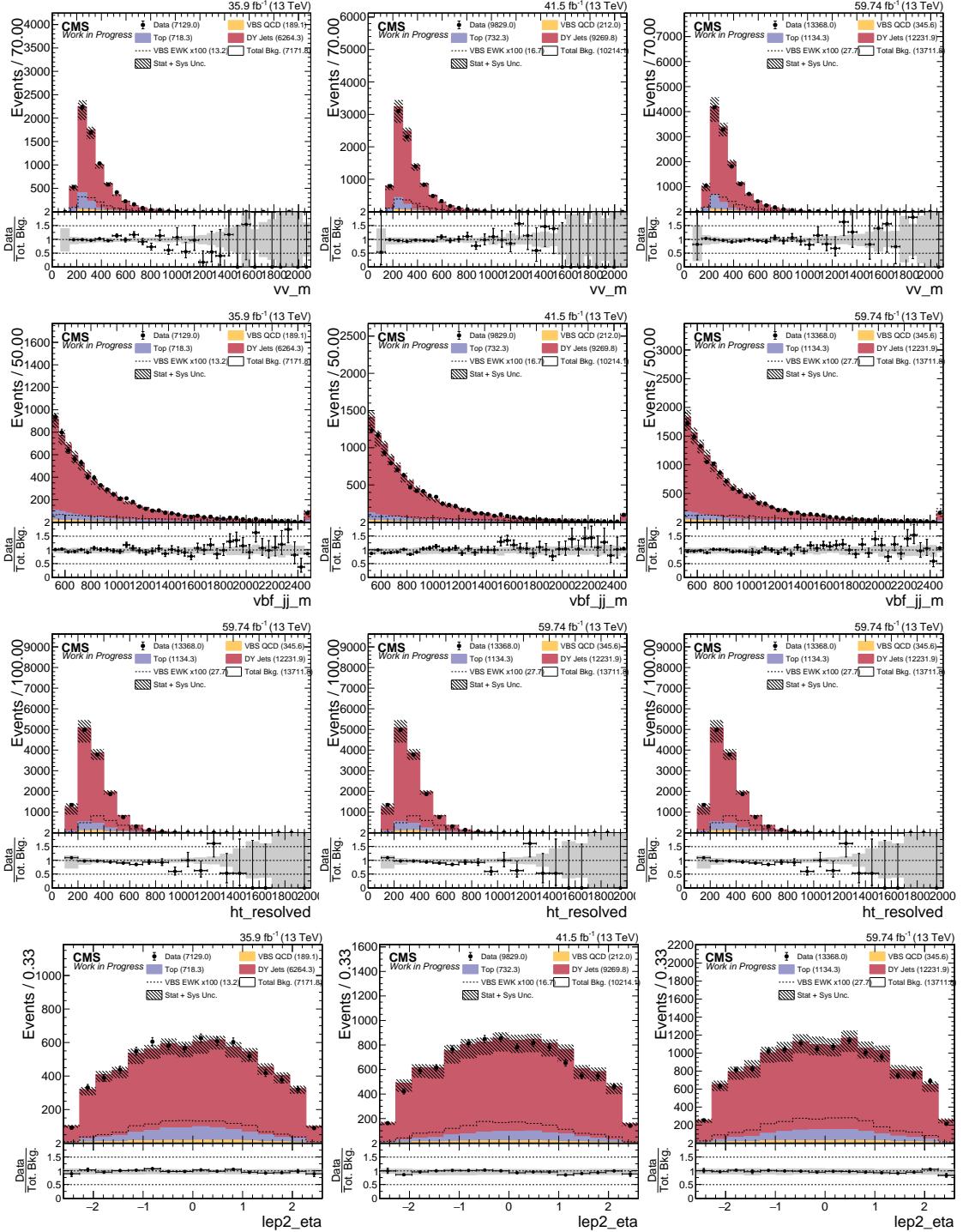


Figure 4.16: Inputs Variables for training Resolved ZV BDT Classifier in signal region. Left to Right: 2016, 2017 and 2018 dataset. Top to Bottom: Diboson invariant mass, VBS tagged jets invariant mass, HT^* (p_T sum of jets), trailing lepton η .

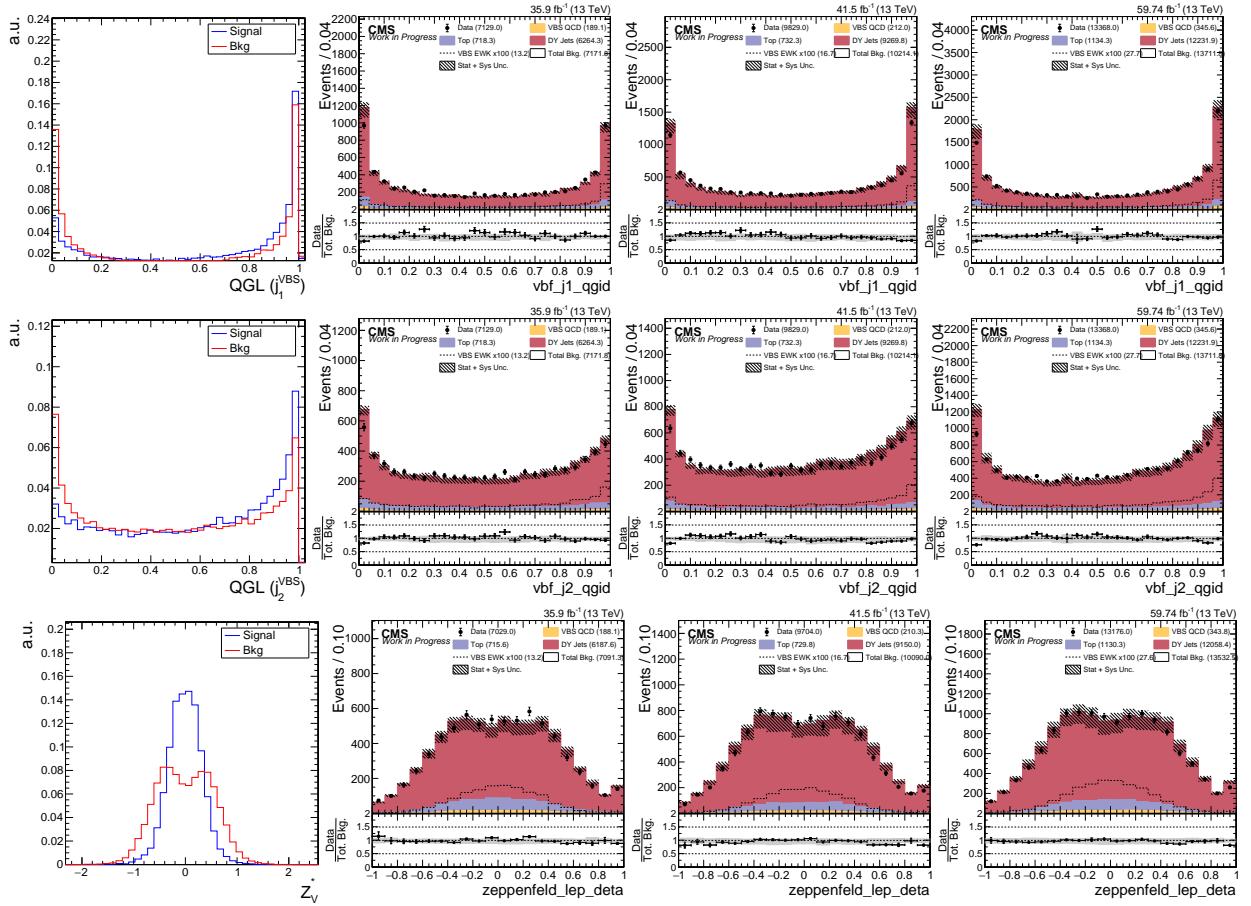


Figure 4.17: Inputs Variables for training Resolved ZV BDT Classifier. Left to Right: Shape comparison of signal and background (combined dataset), control region kinematic distribution in 2016, 2017 and 2018 dataset. Top to Bottom: QGL of leading VBS tagged jet, QGL of trailing VBS tagged jet, Zeppenfeld variable of leptonic boson.

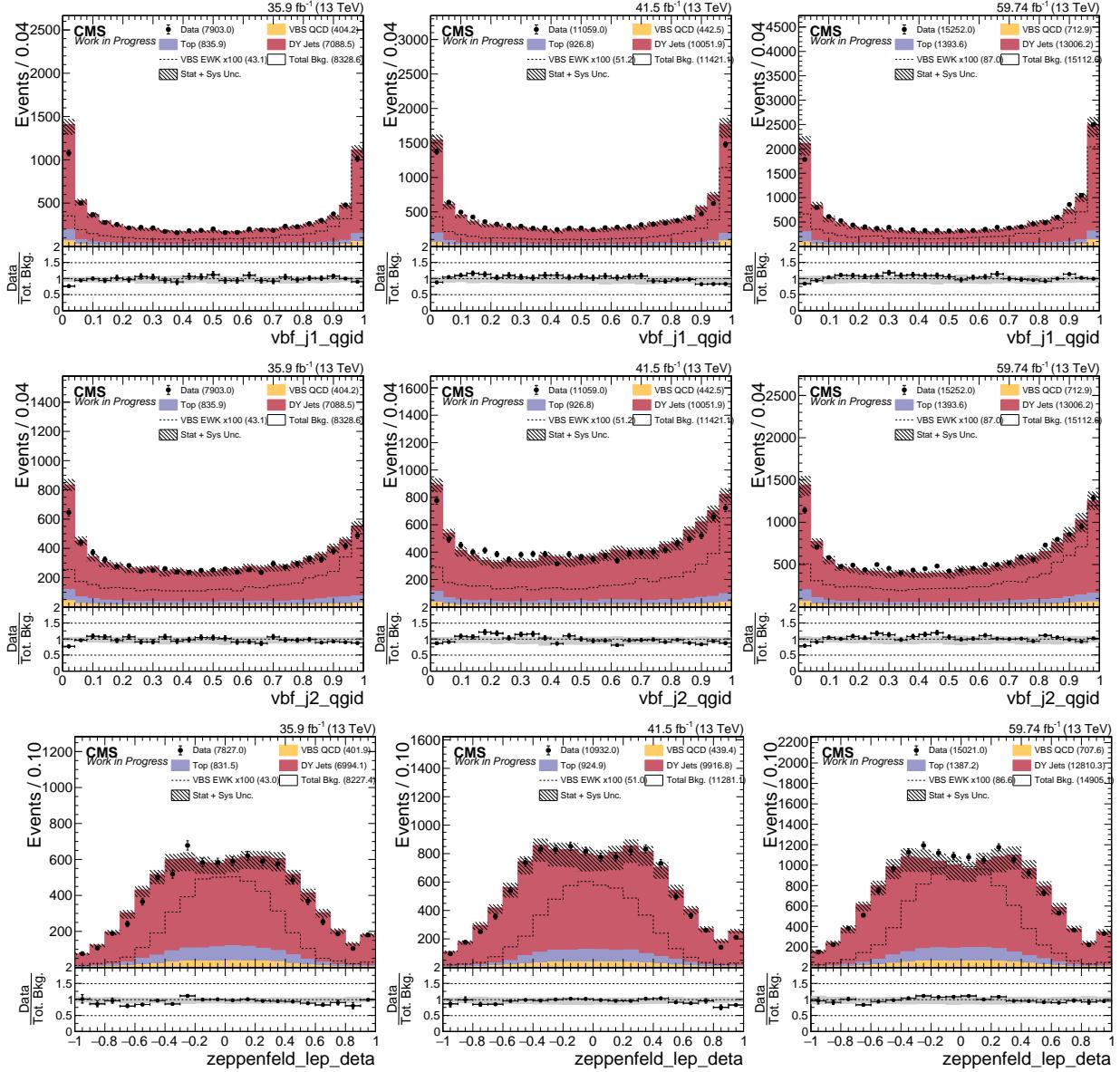


Figure 4.18: Inputs Variables for training Resolved ZV BDT Classifier in signal region. Left to Right: 2016, 2017 and 2018 dataset. Top to Bottom: QGL of leading VBS tagged jet, QGL of trailing VBS tagged jet, Zeppenfeld variable of leptonic boson.

After training, TMVA evaluates BDT input variables and ranks them in terms of importance and separation they provide in the classification (Table 4.7). The correlation matrix of variable is shown in Figure 4.19.

Table 4.7: Training Input Variable Ranking

Channel	Variable	Variable Name	Importance	Separation
Boosted	M_{JJ}^{VBS}	vbf_m	0.2496	0.1348
	Zeppenfeld z^*	zeppLep_deta	0.2396	0.1116
	QGL j_1^{VBS}	vbf2_AK4_qgid	0.1889	0.02413
	QGL j_2^{VBS}	vbf1_AK4_qgid	0.1780	0.02330
	M_{VV}	dibos_m	0.1439	0.005308
Resolved	Zeppenfeld z^*	zeppLep_deta	0.1955	0.1219
	M_{JJ}^{VBS}	vbf_m	0.1822	0.07998
	HT *	ht_resolved	0.1693	0.04201
	QGL j_1^{VBS}	vbf2_AK4_qgid	0.1403	0.02159
	QGL j_2^{VBS}	vbf1_AK4_qgid	0.1341	0.03235
	M_{VV}	dibos_m	0.09098	0.01112
	η_{lep2}	lep2_eta	0.08760	0.01755

The under and over-fitting of trained model is checked by Kolmogorov—Smirnov (K-S) test and ROC curves comparison between training and testing datasets. If the tests are not acceptable, then the training is redone with adjusted parameters. Table 4.8 lists the final training configuration in the TMVA framework for boosted and resolved models. The Figure 4.20 show MVA score and ROC curves of the BDT models. Boosted trained model MVA has 78 % and resolved has 79 % of AUC.

Table 4.8: TMVA configuration of trained models for Boosted and Resolved ZV

TMVA Option	Value	TMVA Option	Value
BoostType	Grad	nCuts	100
NTrees	800	Shrinkage	0.010
MaxDepth	3	UseBaggedBoost	True
MinNodeSize	3%	BaggedSampleFrac	0.6

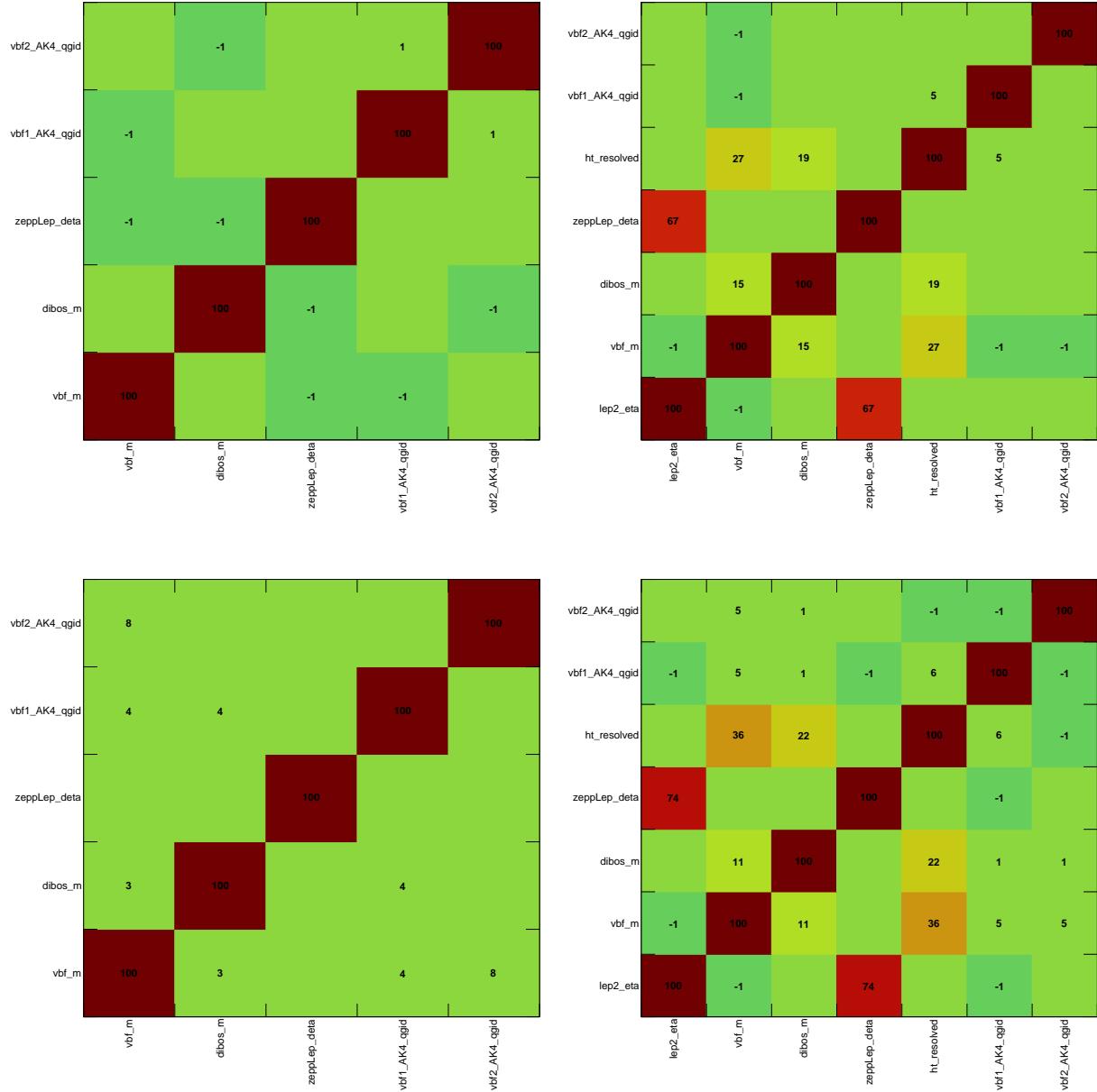


Figure 4.19: Correlation Matrix for Signal and Background. From Left to Right: Boosted, Resolved. From Top to Bottom: Signal, Background

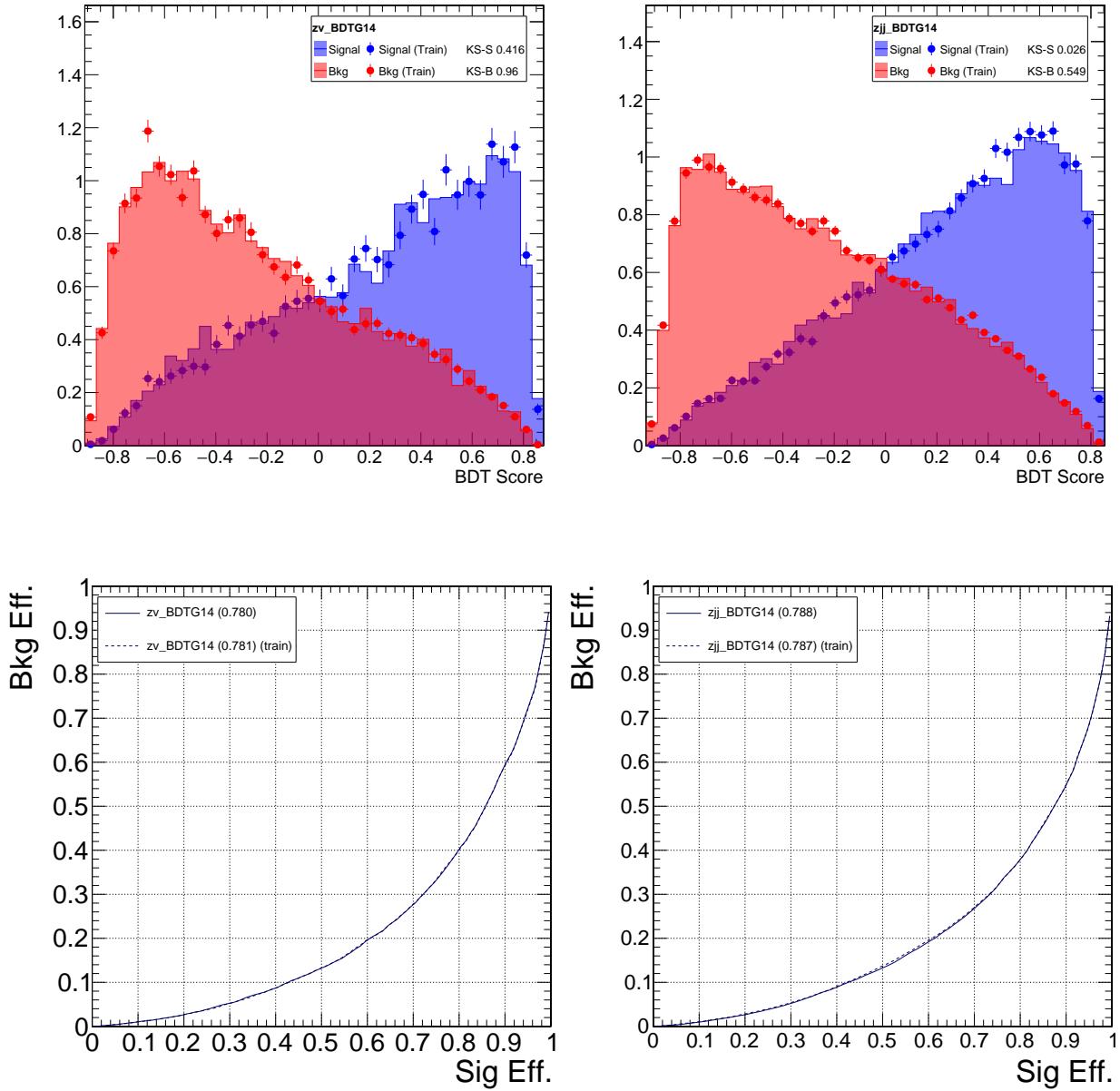


Figure 4.20: From Left to Right: Boosted, Resolved. Top to Bottom: MVA Score of BDT models, ROC Curves.

4.5.3 MVA Score Inference

Using the training models for boosted and resolved ZV, the MVA score is inferred for the complete dataset of data and MC. Then the binning of MVA score is derived such that the signal yield is almost equal in all the bins in signal region. This is done to ensure that we don't introduce a bias in signal sensitivity. MVA score distribution for control region with data is shown in Figure 4.21 and 4.22, and blinded MVA score for signal region is shown in Figure 4.23 and Figure 4.22.

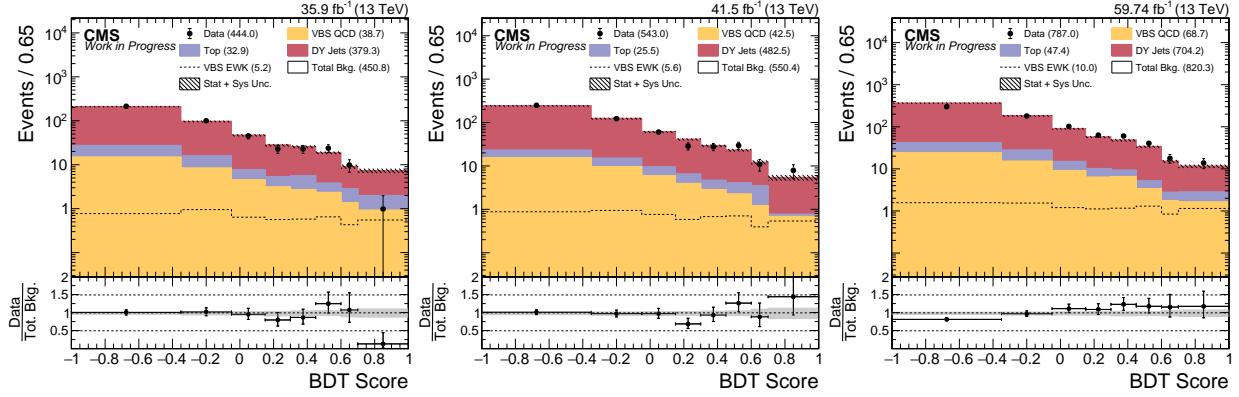


Figure 4.21: MVA Score in Control Region for Boosted ZV Channel.

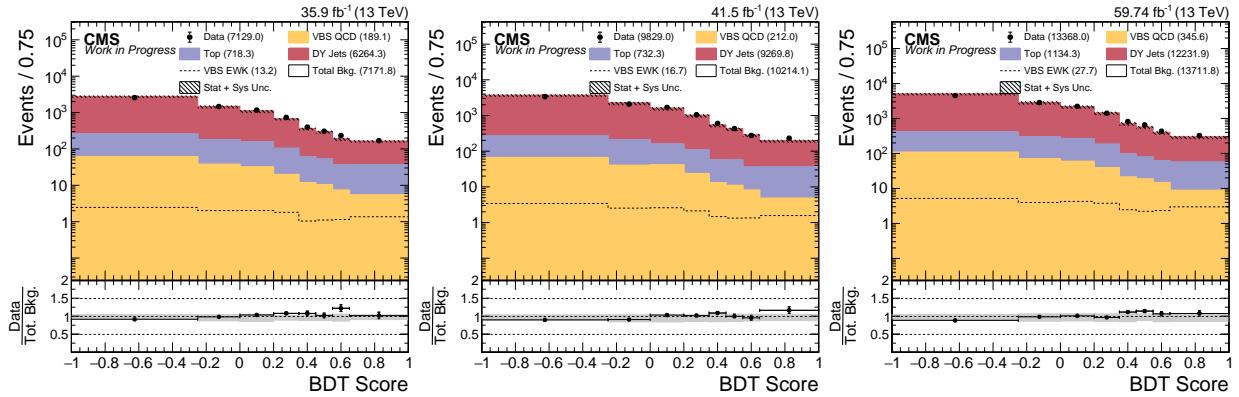


Figure 4.22: MVA Score in Control Region for Resolved ZV Channel.

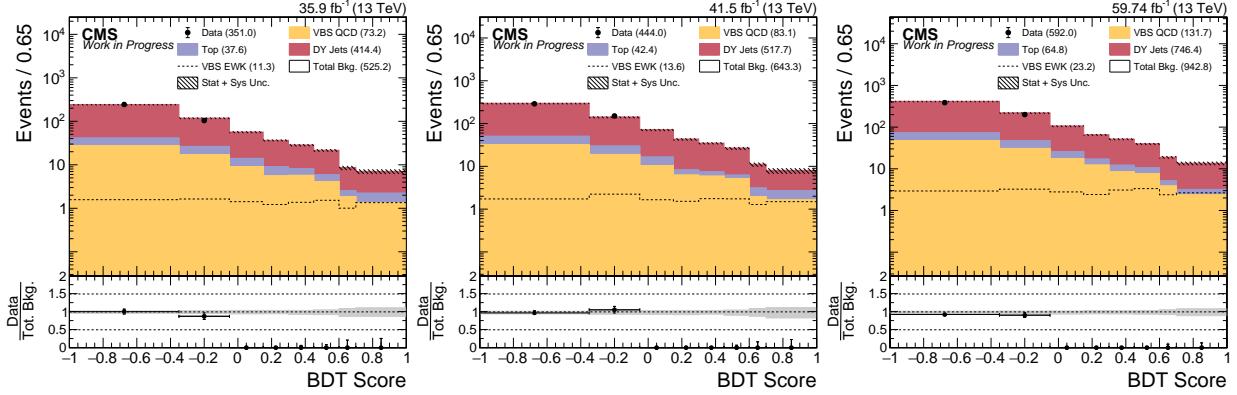


Figure 4.23: MVA Score in Signal Region for Boosted ZV Channel.

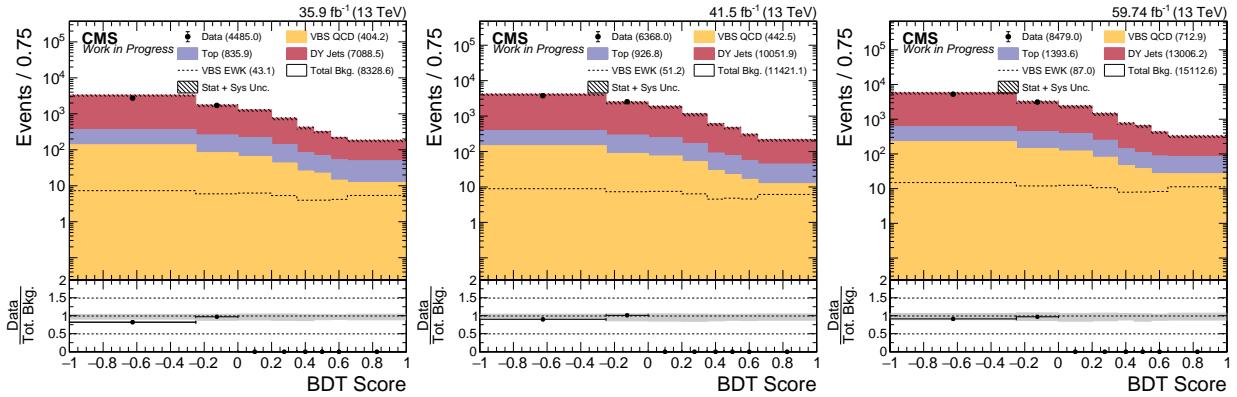


Figure 4.24: MVA Score in Signal Region for Resolved ZV Channel.

4.6 Measurement

To measure our signal, the MVA score in signal region is used as the observable to calculate expected significance and study impact of systematics using impact plots.

`CombineLimit` is command line tool which takes input configuration as plain text file and can perform standard high energy physics statistical analysis. `CombineLimit` is built on `RooFit` and `RooStats` packages which are also part of ROOT package [54, 55].

4.6.1 Statistical and Systematic Uncertainties

MC statistical uncertainties for the bins with more than 10 events are taken into account using “Barlow—Beeston lite” method, and using Poisson probability density function (p.d.f.) otherwise [56].

Systematic uncertainties for correction of the multiplicative type such as integrated luminosity, efficiency, etc., are treated with a log-normal distribution model, i.e. multiplying with factor κ is equivalent to $+1\sigma$ variation and multiplying with $1/\kappa$ to get -1σ variation. For relative uncertainty $\Delta x/x$, the κ can be set as $1 + \Delta x/x$. For example, 2.5% uncertainty with log-normal model on integrated luminosity, κ will be 1.025. Uncertainties set this way are referred to as “flat” systematic uncertainties.

For systematic uncertainties such as JES, the observable used in the measurement is also calculated with JES $+1\sigma$ and -1σ scale factor separately, and the variation up and down value are directly taken from the histogram bin. These types of uncertainties are referred to as “shape” systematic uncertainties.

Following are the systematic uncertainties used in this analysis,

- **Integrated Luminosity:** 2.5% uncertainty on all 2016, 2017, and 2018 year datasets.
- **L1 Prefire Correction:** Shape systematics for 2016 and 2017 year.
- **Lepton Reconstruction and Identification Efficiency:** 1% uncertainty for electron and muons.
- **Lepton momentum scale:** Shape systematics for electron and muon momentum scale, shapes are obtained by varying p_T 3% up and down in selection.
- **JES :** Shape systematics obtained by varying jet p_T with JES up and down, for all the years dataset. The uncertainties are treated separately depending on the source [40].

- **PU Reweighting:** Shape systematics obtained by varying normalization up and down for all the years datasets.
- **Renormalization and factorization scale:** Renormalization and factorization scale uncertainties are calculated for VBS_EWK and VBS_QCD MC datasets, as prescribed by CMS [57] the renormalization and factorization scale factor are varied up and down by the factor of 2, and the envelope of the all the variation is taken as 1σ shape systematics.
- **PDF systematics:** PDF uncertainties are obtained as uniform shape systematic by reweighting MC events cross-section with PDF set.

4.6.2 Significance

Signal significance (Z) is the quantity used to report if there is excess of signal events (s) in presence of background events (b) by testing the ($n = \mu s + b$) hypothesis against the alternate null ($n = b$) hypothesis, where μ is the signal strength.

Signal significance (Z) in the asymptotic limit is calculated using profile likelihood ratio as the test statistics [58],

$$Z = \sqrt{q_0} \quad (4.1)$$

$$q_\mu = -2 \log \frac{L(data|\mu, \hat{\theta}_\mu)}{L(data|\hat{\mu}, \hat{\theta})}, \quad 0 \leq \mu \leq \hat{\mu} \quad (4.2)$$

where $\hat{\mu}$ is the maximum likelihood estimate of μ , and $\hat{\theta}_\mu$ are the nuisance parameters.

As per LHC statistical recommendation [59], likelihood $L(data|\mu, \hat{\theta}_\mu)$ is constructed as,

$$L(data|\mu, \hat{\theta}_\mu) = \text{Poisson}(data|\mu s(\hat{\theta}) + b(\hat{\theta})) \cdot \pi(\hat{\theta}) \quad (4.3)$$

where $\pi(\hat{\theta})$ is the p.d.f. for nuisance parameter.

For calculating expected significance for $\mu = 1$, the observed data is not used, instead n is set equal to total MC events. The expected significance calculated for each category per year dataset, for combined dataset per category and for combined dataset and category is listed in Table 4.9.

Table 4.9: Expected Signal Significance

Channel	2016	2017	2018	Combined
Boosted	0.64	0.66	0.95	1.33
Resolved	0.40	0.37	0.56	0.78
Combined				1.52

4.6.3 Impact Plots

Impact plots are way to visualize which nuisance parameter θ have largest effect on parameter of interest μ and measure correlation from the direction of $+1\sigma$ (correlated) and -1σ (anti-correlated). Figure 4.25, 4.26 shows impacts for top 60 nuisance parameter, and Figure B.8, B.9 in the Appendix B.3 for the rest of nuisance parameter.

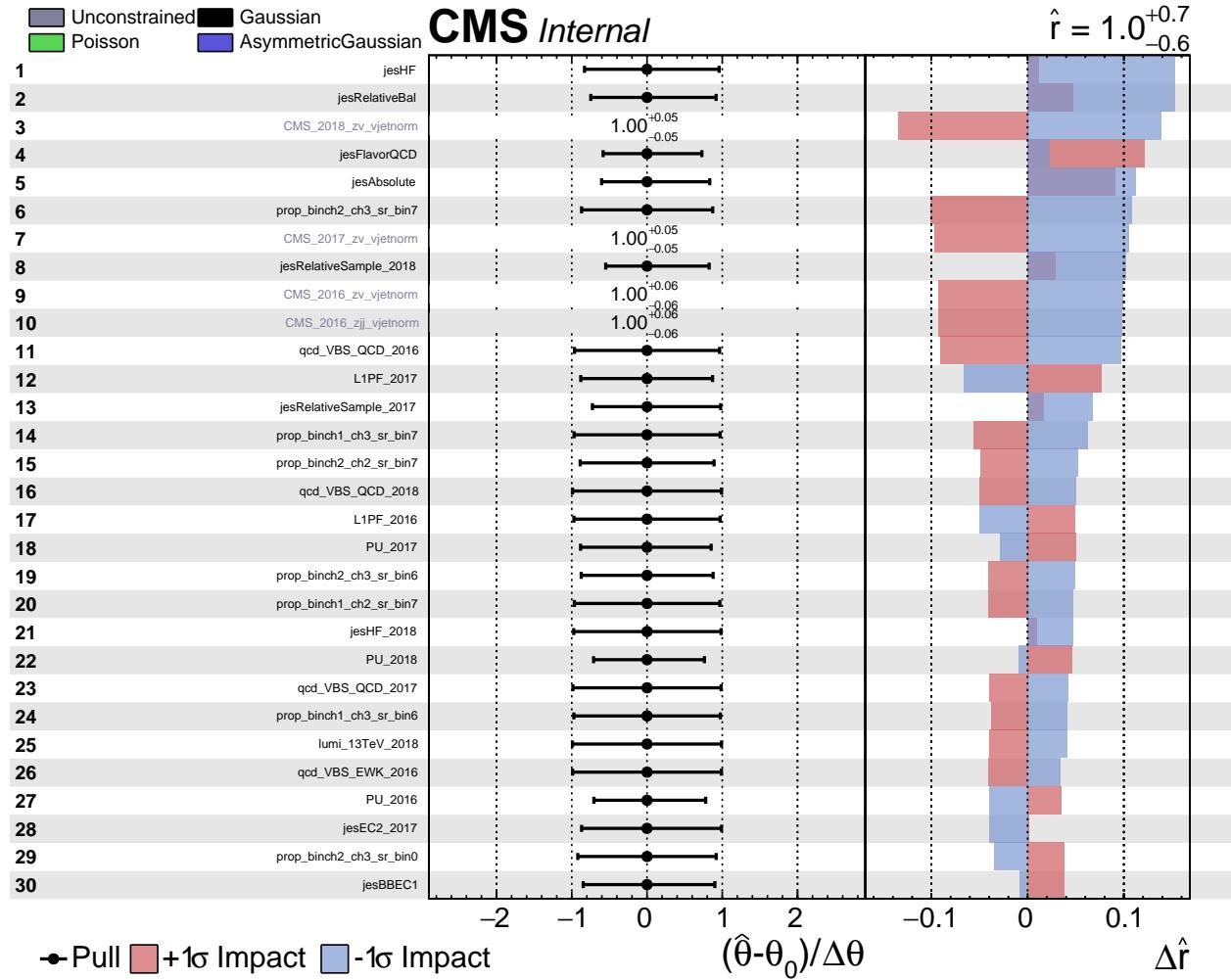


Figure 4.25: Impact Plots of nuisance parameters from 1 to 30.

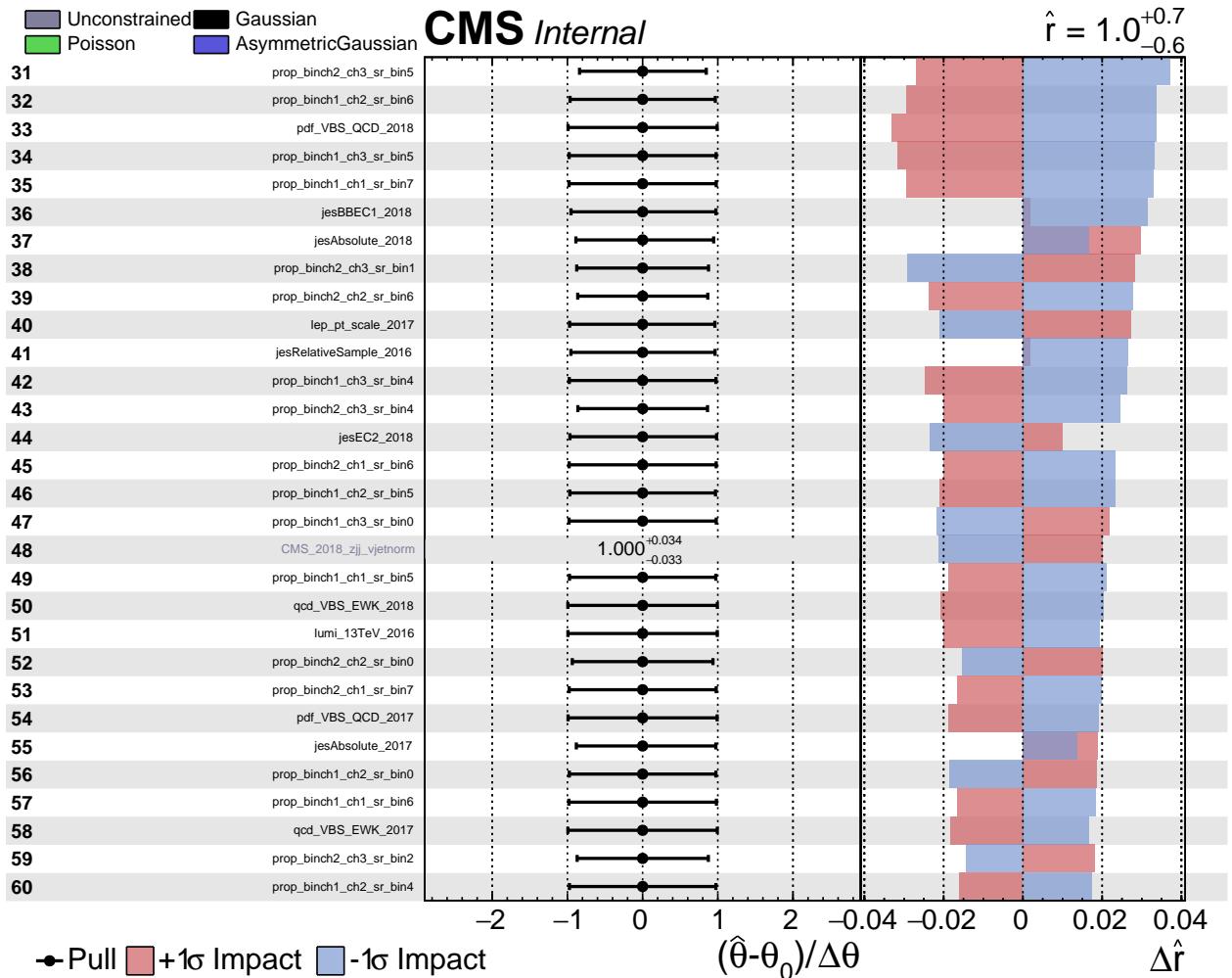


Figure 4.26: Impact Plots of nuisance parameters from 31 to 60.

4.6.4 Postfit Plots

Since the analysis is currently blinded in signal region. The Asimov dataset is used in signal region i.e. data is set equal to sum of expected signal and background events. Postfit distribution in signal region using Asimov dataset for Boosted ZV are shown in Figure 4.21 and for Resolved ZV are shown in Figure 4.28. MC processes overall prefit and postfit (signal + background and background-only) values in control and signal region are listed in Table 4.10 for Boosted ZV and in Table 4.11 for Resolved ZV category.

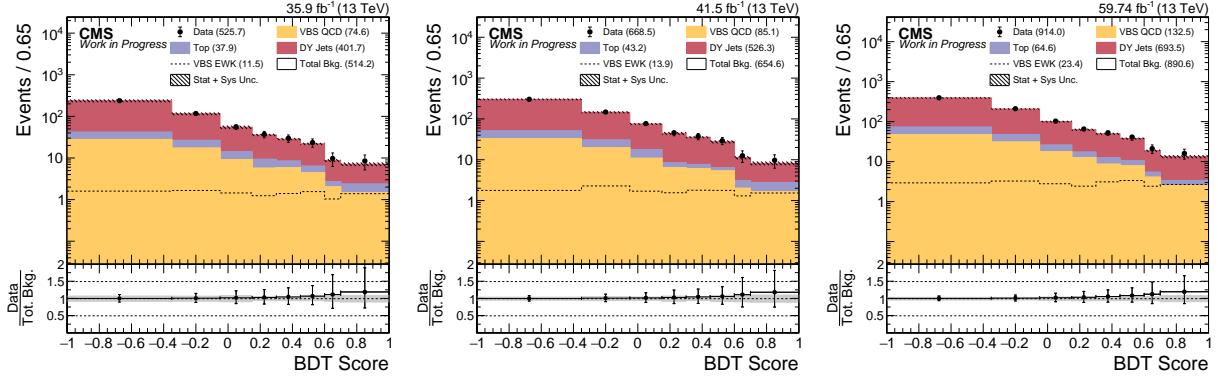


Figure 4.27: (Asimov Data) MVA Score postfit in Signal Region for Boosted ZV Channel.

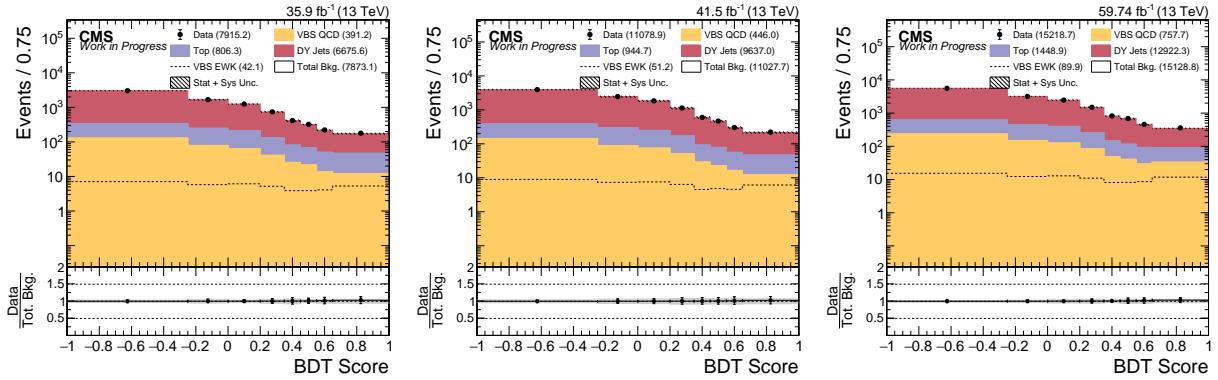


Figure 4.28: (Asimov Data) MVA Score postfit in Signal Region for Resolved ZV Channel. (Asimov Data)

Table 4.10: Prefit and postfit yields of MC processes

Region	Process	Pre-fit	S+B Fit	B-Only Fit
Boosted ZV 2016				
CR	DY+Jets	379.258 ± 11.645	366.292 ± 20.823	372.397 ± 17.543
CR	Top	32.874 ± 1.214	32.936 ± 1.146	32.949 ± 1.255
CR	VBS_EWK	5.230 ± 0.156	5.239 ± 8.585	0.000 ± 0.000
CR	VBS_QCD	38.659 ± 1.250	38.730 ± 1.199	38.748 ± 1.286
SR	DY+Jets	414.424 ± 37.150	401.658 ± 33.875	410.467 ± 33.476
SR	Top	37.557 ± 7.538	37.905 ± 7.583	38.173 ± 7.722
SR	VBS_EWK	11.280 ± 1.097	11.499 ± 19.267	0.000 ± 0.000
SR	VBS_QCD	73.244 ± 5.459	74.605 ± 4.790	75.410 ± 4.942
Boosted ZV 2017				
CR	DY+Jets	482.465 ± 14.680	477.902 ± 22.557	484.890 ± 18.593
CR	Top	25.458 ± 1.980	25.347 ± 1.839	25.335 ± 1.986
CR	VBS_EWK	5.584 ± 0.167	5.593 ± 8.960	0.000 ± 0.000
CR	VBS_QCD	42.475 ± 1.295	42.524 ± 1.262	42.543 ± 1.333
SR	DY+Jets	517.728 ± 28.192	526.307 ± 29.981	537.022 ± 21.508
SR	Top	42.445 ± 5.199	43.179 ± 5.010	43.436 ± 5.155
SR	VBS_EWK	13.635 ± 1.135	13.918 ± 22.190	0.000 ± 0.000
SR	VBS_QCD	83.087 ± 5.566	85.124 ± 5.179	86.000 ± 5.310
Boosted ZV 2018				
CR	DY+Jets	704.184 ± 21.318	658.283 ± 26.685	670.205 ± 23.886
CR	Top	47.431 ± 1.545	47.601 ± 1.435	47.630 ± 1.489
CR	VBS_EWK	10.045 ± 0.296	10.078 ± 10.857	0.000 ± 0.000
CR	VBS_QCD	68.708 ± 2.167	68.950 ± 1.982	68.996 ± 2.065
SR	DY+Jets	746.410 ± 41.307	693.492 ± 29.658	711.877 ± 26.629
SR	Top	64.774 ± 3.933	64.638 ± 2.715	65.198 ± 3.368
SR	VBS_EWK	23.232 ± 1.294	23.391 ± 24.686	0.000 ± 0.000
SR	VBS_QCD	131.658 ± 7.743	132.474 ± 6.005	133.943 ± 6.877

Table 4.11: Prefit and postfit yields of MC processes

Region	Process	Pre-fit	S+B Fit	B-Only Fit
Resolved ZV 2016				
CR	DY+Jets	6264.349 ± 190.352	6149.237 ± 92.197	6162.392 ± 85.695
CR	Top	718.306 ± 21.460	734.229 ± 21.392	735.064 ± 21.753
CR	VBS_EWK	13.240 ± 0.395	13.527 ± 29.811	0.000 ± 0.000
CR	VBS_QCD	189.095 ± 5.645	193.259 ± 5.631	193.479 ± 5.720
SR	DY+Jets	7088.493 ± 633.668	6675.562 ± 330.385	6710.664 ± 266.921
SR	Top	835.933 ± 69.920	806.328 ± 40.673	809.491 ± 39.805
SR	VBS_EWK	43.133 ± 2.811	42.092 ± 91.329	0.000 ± 0.000
SR	VBS_QCD	404.160 ± 34.797	391.192 ± 19.933	392.868 ± 20.162
Resolved ZV 2017				
CR	DY+Jets	9269.820 ± 280.782	8831.423 ± 109.274	8846.941 ± 100.456
CR	Top	732.264 ± 22.326	738.436 ± 21.576	739.105 ± 22.234
CR	VBS_EWK	16.719 ± 0.499	16.839 ± 39.897	0.000 ± 0.000
CR	VBS_QCD	212.000 ± 6.326	213.551 ± 6.143	213.741 ± 6.281
SR	DY+Jets	10051.850 ± 952.129	9637.035 ± 645.399	9681.931 ± 511.040
SR	Top	926.759 ± 64.180	944.658 ± 45.063	948.703 ± 44.739
SR	VBS_EWK	51.194 ± 3.178	51.201 ± 118.771	0.000 ± 0.000
SR	VBS_QCD	442.505 ± 34.337	445.977 ± 23.882	447.682 ± 24.279
Resolved ZV 2018				
CR	DY+Jets	12231.940 ± 370.848	11810.172 ± 129.492	11836.949 ± 122.098
CR	Top	1134.261 ± 33.468	1169.084 ± 28.849	1170.880 ± 31.799
CR	VBS_EWK	27.689 ± 0.816	28.538 ± 48.075	0.000 ± 0.000
CR	VBS_QCD	345.569 ± 10.189	356.174 ± 8.782	356.722 ± 9.687
SR	DY+Jets	13006.154 ± 1365.518	12922.276 ± 541.237	12998.396 ± 527.575
SR	Top	1393.580 ± 115.044	1448.865 ± 68.813	1455.162 ± 65.230
SR	VBS_EWK	87.014 ± 5.266	89.902 ± 149.233	0.000 ± 0.000
SR	VBS_QCD	712.871 ± 58.615	757.654 ± 35.477	762.648 ± 32.311

4.7 Results Discussion and Future Prospects

Using multivariate analysis expected signal significance of 1.5σ is reported. The analysis is currently performed blinded and the full results are expected in early next year 2023. This is a small piece of overall VBS efforts at CMS. Ultimately the plan is to search for VBS in all the diboson channels and all the final states.

Run3 (2022–2025) of LHC proton-proton collisions is expected to collect an integrated luminosity of 250 fb^{-1} . This is approximately double the combined luminosity of Run1 and Run2. We expect the larger dataset and the usage of advanced multivariate techniques will enable us to provide evidence for VBS in semileptonic ZV channel.

Additionally, constraints on the structure of quartic vector boson interactions in the framework of dimension-8 Effective Field Theory (EFT) operators are in progress using this channel and the semileptonic WV channel as well.

Jets at large η are key to VBS measurement. The planned upgrade of the endcap calorimeter with High Granularity Calorimeter (HGCAL) will improve our capabilities to identify these jets and reject pileup contamination.

CHAPTER 5

HIGH GRANULARITY CALORIMETER UPGRADE

By the start of the Run4 period of proton-proton collisions at LHC, the collision energy is expected to reach its full design limit of 14 TeV and commissioning of High Luminosity LHC (HL-LHC) is expected to increase the luminosity by 10 times. The integrated luminosity collected by the end of Run5 (2038) is expected to be 3000 fb^{-1} .

With increased luminosity, the CMS detector will get higher dose of radiation and the average number of pileup interactions will be of the order $O(140)$. The endcap calorimeters ECAL and HCAL will suffer irreparable damage due to the much higher radiation dose received from the increased luminosity in those regions. The High Granularity Calorimeter (HGCAL) is an upgrade that will replace current endcap calorimeters (ECAL and HCAL). HGCAL is expected to be completed and installed during Long Shutdown 3 (LS3) (2026–2028).

This chapter will discuss broadly the design of HGCAL, especially its scintillator section with studies done at Northern Illinois Center for Accelerator and Detector Development (NICADD) towards its upgrade. For complete design see the Technical Design Report (TDR) [60].

5.1 Technical Design and Requirements

As mentioned, the HGCAL will replace current endcap ECAL and HCAL. Figure 5.1 shows exactly where the new detector will be placed. The image on the right in Figure 5.1

shows a side view ($z - r$ plane) of the detector. Starting from the left i.e. innermost layers is the Calorimeter Endcap Electromagnetic (CE-E) whose active layers are made of silicon cells ($\approx 0.5 - 1 \text{ cm}^2$). The majority of the detector in longitudinal depth is Calorimeter Endcap Hadronic (CE-H) whose starting few active layers are also all silicon followed by mixed layers with silicon cells in lower rings of the layer and scintillator tiles ($\approx 5 - 31 \text{ cm}^2$) in the rest. Silicon is radiation hard material i.e. the response of silicon cells will be adequate under accumulated radiation dose over time. But silicon is also expensive and for such high coverage area in HGCAL, it significantly increases the construction cost of the project. For this reason scintillator tiles are used wherever radiation doses are low enough as they offer good performance at reasonable cost. Some of the main reasons for high cell count, lateral and longitudinal granularity are to preserve energy resolution after 3000 fb^{-1} , aid particle flow (PF) reconstruction in rejecting energy deposit from pileup using precise timing measurement, and being able to observe narrower jets with $R = 0.2$. To be able to deliver these requirements both silicon cells and scintillator tiles need to have good signal-to-noise ratio (S/N) even after 3000 fb^{-1} .

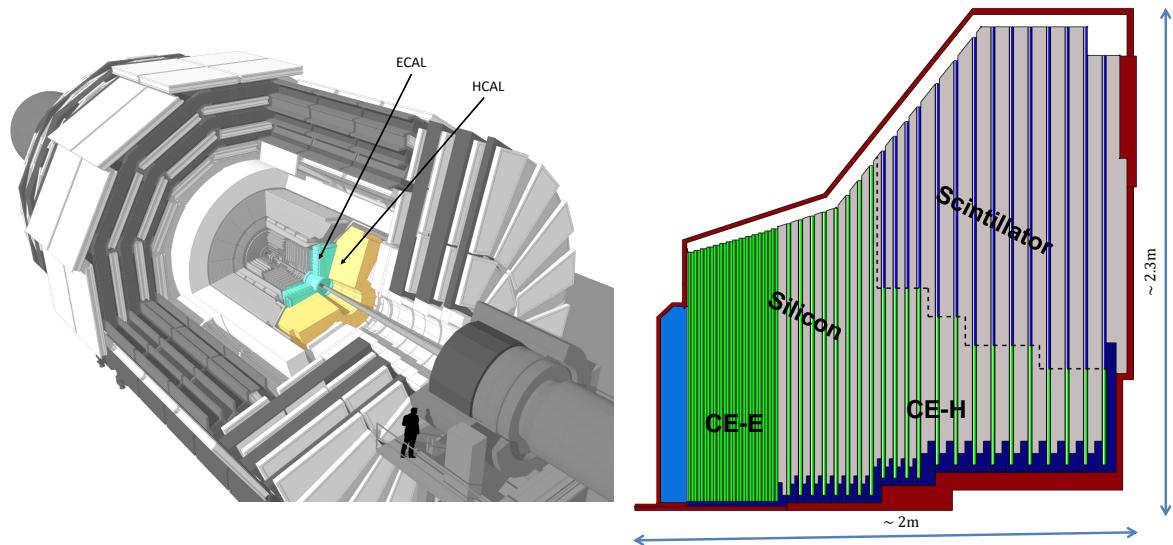


Figure 5.1: Overview of HGCAL position and quadrant view of Calorimeter Endcap (CE). Left image shows current endcap ECAL and HCAL highlighted. Right image shows quadrant view of HGCAL from the side [61, 62].

5.2 Scintillator Tiles and SiPMs

Silicon cells will be directly fabricated 8 inch silicon wafers (432 cells) (shown in Figure 5.2 as yellow and green colored hexagons), whereas each scintillator tile will need to be prepared separately and assembled in the form of a “tileboard” (about 8×8 tiles). Figure 5.2 shows scintillator tile boundary with red grid lines. There are 288 scintillator tiles in each ring, and each layer has different number of rings of scintillator tiles with a maximum of 42 rings. To reduce the production cost and assembly complexities, scintillator sizes are same for every two rings. The ring number is used to identify tile size, for example R18–19 (Figure 5.4) is the size of tiles in ring number 18 and 19.

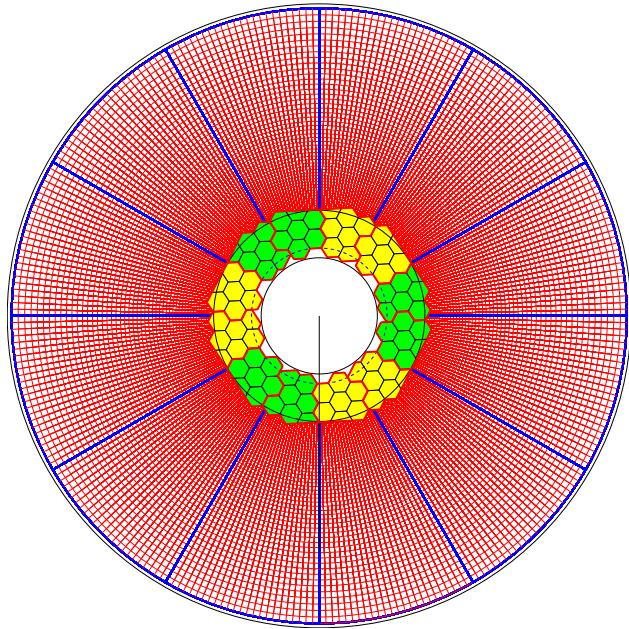


Figure 5.2: CE-H mixed layer 22. This is the last layer in CE-H (Right image Figure 5.1). Yellow and green hexagons are the 8 inch silicon modules and the red grid lines shows 11,520 (40 rings) scintillator tiles. [60].

5.2.1 Scintillator Materials

Materials which scintillate i.e. emits light whenever ionizing radiation passes through it are known as scintillators. Scintillators can be crystals (current barrel and endcap ECAL), liquid or plastic and are broadly divided into two categories, organic or inorganic. HGCAL will make use of organic plastic scintillator. The most commonly used base material in scintillators of this type are polyvinyltoluene (PVT) and polystyrene (PS). PVT and PS are base material and by themselves cannot serve as scintillators for a couple of reasons. The light emitted is at lower wavelengths (ultraviolet) and they are barely transparent to this light. For these reasons they have a small percentage of dopants added which absorb the scintillation light and re-emit it at larger wavelengths (visible). A second dopant is added to further increase the attenuation length of the light emitted, so that light can be collected and coupled to a light detection system.

The plastic scintillator can be produced via extrusion, casting or injection molding. PVT and PS are used as base material for the cast and injection molded scintillator tiles, respectively.

Cast scintillators are generally brighter (in terms of light output) than injection molded scintillators, but the production cost of cast scintillator per unit area is generally higher and additional machining costs. Choice of scintillator in addition to silicon photomultiplier (SiPM), which will be discussed next, impacts the cost-performance optimization.

5.2.2 SiPMs

SiPM is a solid-state device that converts incident photons to electric current with a large gain (10^5 to 10^6). SiPMs achieve this with pixels ($10\text{ }\mu\text{m}$ to $100\text{ }\mu\text{m}$ in size) connected

in parallel, where each pixel is an avalanche photodiode (APD) combined with quenching resistor. For example a SiPM of active area 2 mm^2 with $15\text{ }\mu\text{m}$ pixel has approximately 9,000 pixels. Commercially SiPMs produced by Hamamatsu Photonics are also known as Multi-Pixel Photon Counter (MPPC) [63]. Figure 5.3 shows SiPM next to tip of a pen.

SiPM operates in reverse bias (Geiger Mode) with operating voltage (V_0) above breakdown (V_{BR}). It has a linear relationship between $V (= V_0 - V_{BR})$ and the gain. In addition to low operating voltage ($40 - 60$ V) the power consumption is also low. For HGCAL Hamamatsu S14160 [64] is being considered with active area $2, 4$ and 9 mm^2 . Larger area means larger signal, but also higher intrinsic noise and power consumption.

For HGCAL the SiPMs will be mounted on printed circuit board (PCB), and prepared scintillator tiles will be directly glued over the SiPMs with the dimple centered on the SiPM's active area [65].



Figure 5.3: SiPM

5.2.3 Scintillator Tiles

Scintillator tiles coupled directly to SiPMs alone cannot provide sufficient response to the minimum ionizing particles (MIPs). Additionally the response will be non-uniform at the SiPM. Higher signal and uniformity of response, can be achieved by either coating or wrapping the tile in reflective material [66]. Enhanced Specular Reflector (ESR) is a multi-layer highly reflective material with $65\text{ }\mu\text{m}$ thickness (Figure 5.5), is the material chosen for wrapping the scintillator tiles.

To wrap 239,616 individual tiles with ESR is a very challenging task. At NIU we have built an automated wrapping machine (Section 5.3) to wrap scintillator tiles with speed and repeatability.

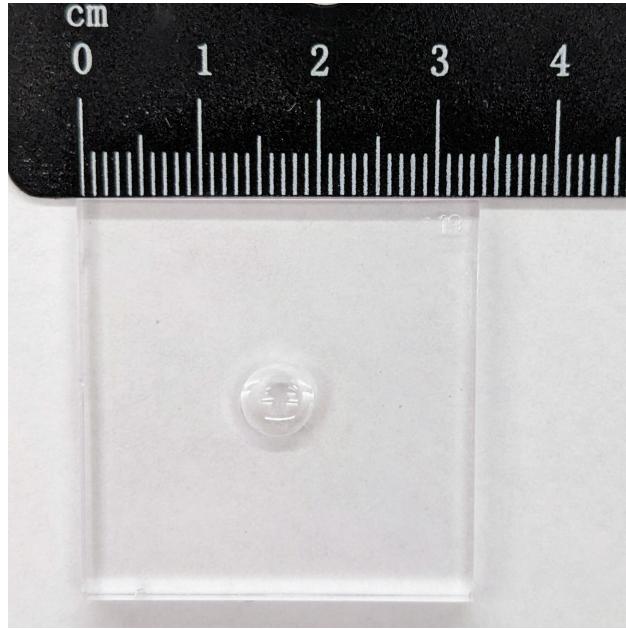


Figure 5.4: Scintillator tile with dimple

The final wrapped tile in complete automated process with wrapping machine built at NIU is shown in Figure 5.6.

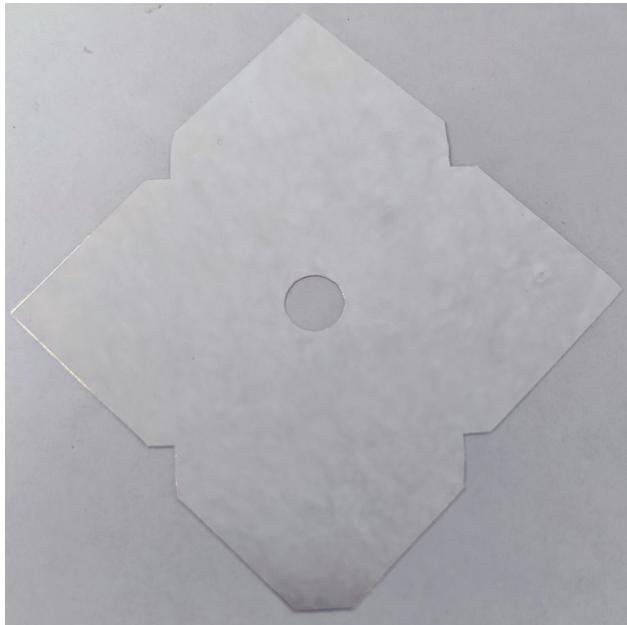


Figure 5.5: ESR wrapper cut for tile size R18–19

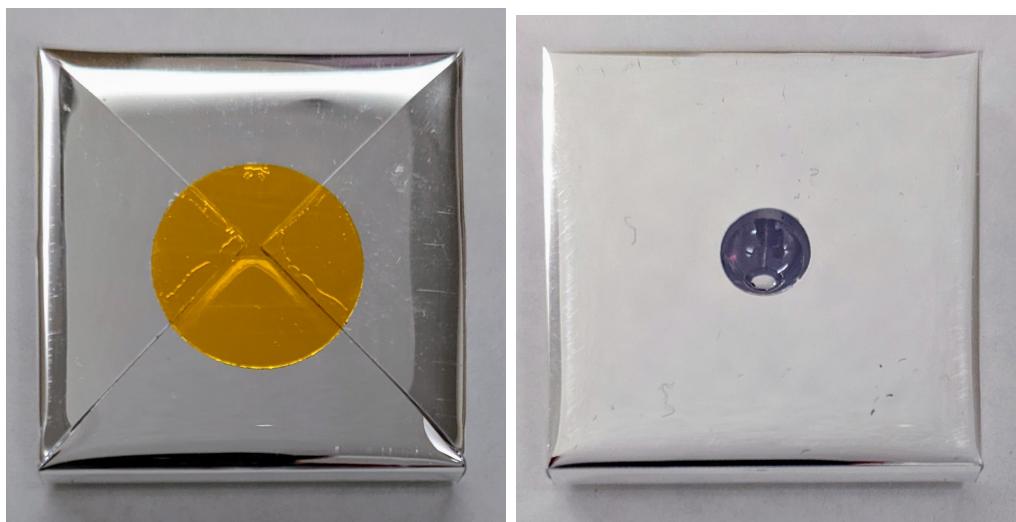


Figure 5.6: R18–19 scintillator tile wrapped in ESR. Left: Upside of the tile, with Kapton sticker holding the flaps. Right: Bottom side with center hole over dimple.

5.3 Automated Wrapping of Scintillator Tiles

As mentioned in Section 5.2.3, the number of wrapped scintillator tiles required for the HGCAL is very large, 239,616 tiles. In addition to the large number of tiles, the tiles will be of different sizes and repeatability in wrapped tile quality is required for reliable performance of the detector. Figure 5.7 and 5.8 shows the overview of the automated wrapping machine being developed at NIU in collaboration with Deutsches Elektronen-Synchrotron (DESY).

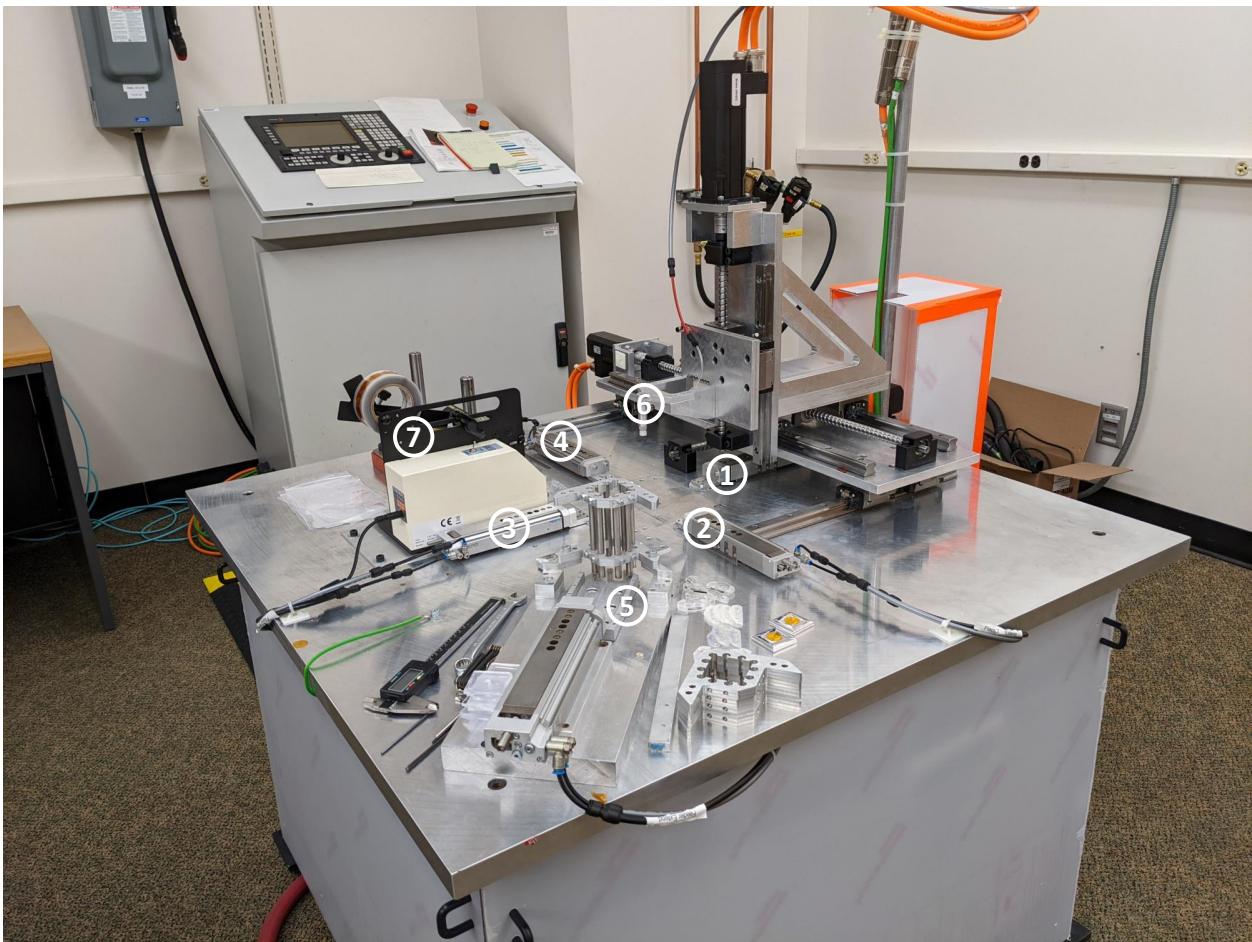


Figure 5.7: Automated scintillator tile wrapper overview. Shown in picture with labels are, 1–4: Actuator arms for folding the cut ESR flaps over the tiles, 5: Tile magazine and dispenser assembly, 6: z-axis (vertical up/down) end-effector with vacuum suction, 7: Kapton sticker dispenser.

The machine is built to provide precise motion in *xyz*-axes and controlled by a computer numerical control (CNC) using G-code programming language. One of the key components of the wrapper machine are actuators which have arms that can extend and retract quickly with the pressurized air. The wrapper uses both pressurized air and vacuum which are controlled by a solenoid switch using a programmable logic controller running in G-code programs.

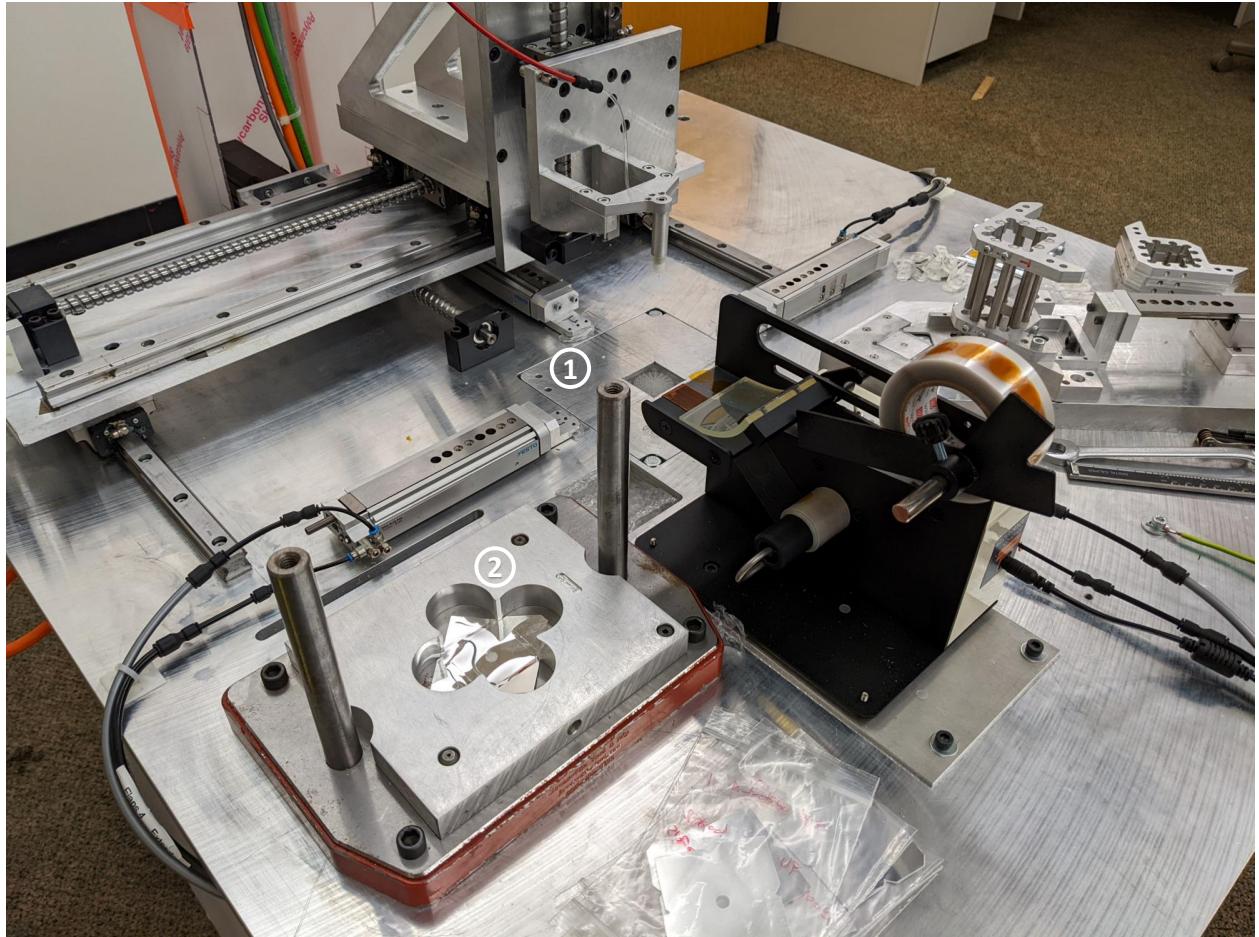


Figure 5.8: Tile folding station and ESR magazine. Shown in picture with labels 1: tile folding station, and 2: ESR magazine.

The final wrapped tile as shown in Figure 5.6 is done by the machine in following steps:

- **ESR Pickup, and Placement:** z -axis (vertical up/down) end-effector picks up the cut ESR using vacuum suction from the ESR magazine, and then places it precisely over the tile pocket in the folding station. Placed ESR over tile pocket is held into its place with the help of vacuum suction, which is built into the tile pocket seat.
- **Tile Dispenser, Pickup, and Placement:** Tiles will be stacked in the magazine at the tile dispenser location and an actuator arm will push out a tile completely and hold in place (right image in Figure 5.9) until the z -axis end-effector makes contact and vacuum suction is turned on. Now, the actuator arm is retracted, and tile is picked up and placed over ESR at the folding station.
- **ESR Flaps Folding:** After ESR and tile have been placed, the tile pocket retracts with the help of actuator connected to it from the bottom. At this stage tile is inside pocket and flush with top of the table. ESR flaps are out and vertical. Now the four actuator arms as shown in Figure 5.10 extend and close ESR flaps.
- **Sticker Application:** The folding actuator arms stay in place until sticker is applied using z -axis end-effector which picks up the sticker from sticker dispenser (left image Figure 5.9) and applies it on the center of the tile where all flap corners meet.
- **Wrapped Tile:** Final step in wrapping is retracting the folding actuator arms, pushing out the tile and turning off the vacuum suction of the tile pocket. Now z -axis end-effector picks up the wrapped tile and drops into the collection basket. Wrapped tile in Figure 5.6 is wrapped automatically with this procedure.

A supplemental video showing wrapping of scintillator tile in ESR film by the machine is provided [67].

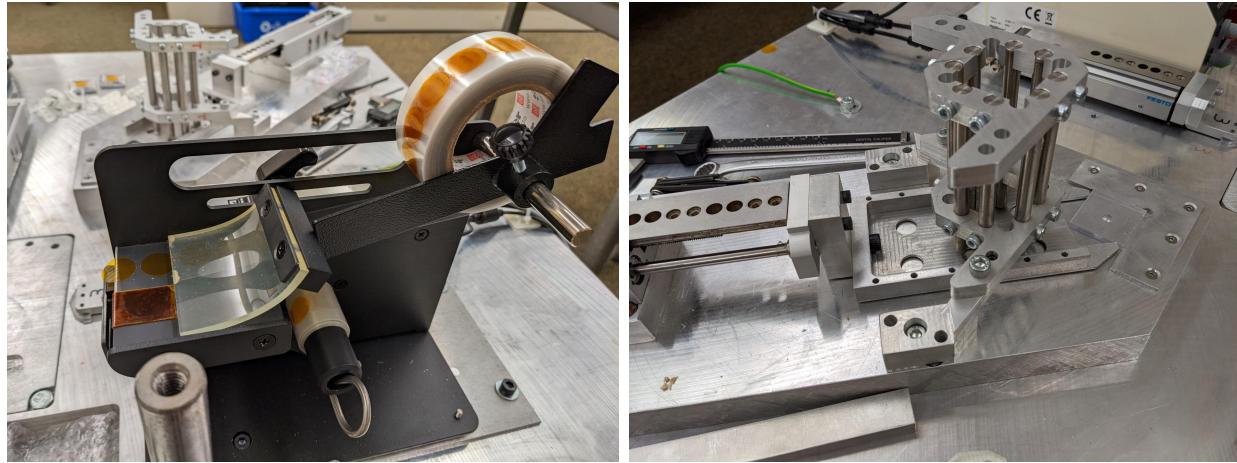


Figure 5.9: Sticker and tile dispenser, Left: Sticker dispenser with kapton tape roll. Right: Tile dispenser with tile magazine.

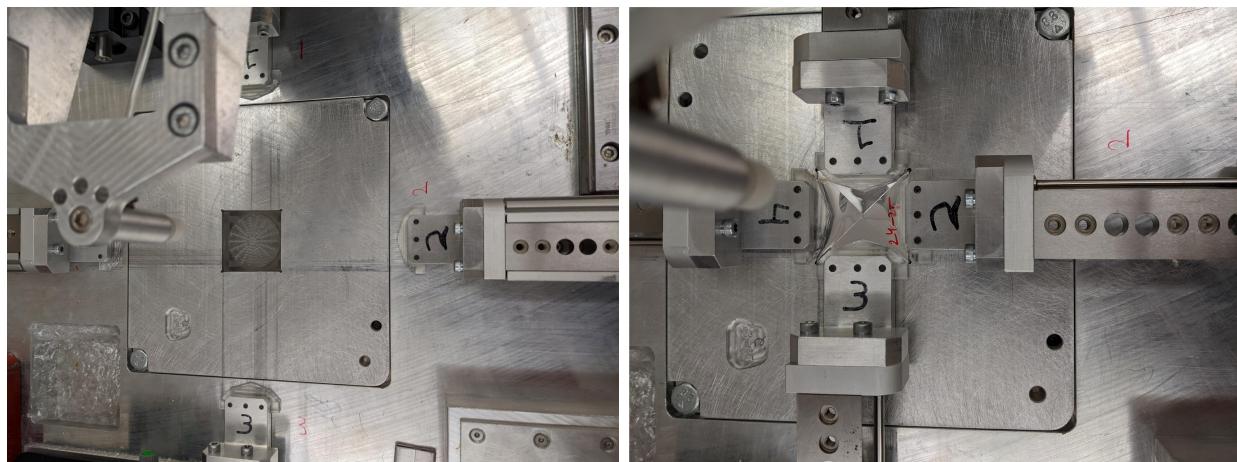


Figure 5.10: Tile folding station. Left: Flushed tile pocket at the center and retracted actuators. Right: After all the arms are extended during wrapping.

5.4 Signal-to-Noise Ratio

As discussed earlier in the Section 5.1, for the HGCAL to retain its performance till the end-of-life it needs good S/N , which is defined as $S/N > 3$. In this section we will discuss formulation and inputs to the S/N calculation, and the results of optimal configuration by minimizing cost and retaining good S/N .

5.4.1 Formulation

Signal-to-Noise formulation is based detecting minimum ionizing particles (MIPs). For the scintillator tile coupled directly with SiPM is formulated as [68]:

$$\frac{S}{N} = \frac{(\text{MIP}_{\text{response}}) \sqrt{\frac{A_{t,\text{ref}}}{A_t}} \left(\frac{A_s}{A_{s,\text{ref}}} \right) (\text{Radiation Loss})}{(\text{SiPM}_{\text{noise,base}}) \sqrt{\frac{A_s}{A_{s,\text{base}}}} \sqrt{1.88^{\frac{T_s - T_{s,\text{base}}}{10^\circ \text{C}}}} \sqrt{\frac{f}{f_{\text{base}}}}} \quad (5.1)$$

where,

- $\text{MIP}_{\text{response}}$: measured in photo electrons (p.e.) is the response of the scintillator tile light yield with a scale factor to account for SiPM device photon detection efficacy (PDE) difference used during testbeam measurement and SiPM expected to be used.
- A_t : is the area of the tile for which the S/N is being evaluated and subscript ref means the area of tile used in the testbeam MIP measurement.
- A_s : the active area of the SiPM device coupled to scintillator tile.

- **Radiation Loss:** is the loss in light output of the scintillator due to the radiation dose received, it is expressed as:

$$e^{-R/D_c} \quad (5.2)$$

$$D_c = (6.0 \text{ Mrad}) \left(\frac{R}{1 \text{ krad/hr}} \right)^{0.35} \quad (5.3)$$

where R is the dose rate in krad/hr, and D_c is the dose constant, and both are obtained from FLUKA simulations.

- $\text{SiPM}_{\text{noise,base}}$: is the root means square (RMS) value in p.e. of intrinsic noise of SiPMs from thermal excitations (also called dark current rate (DCR)), irradiation of silicon also increases this noise.
- $T_{s,\text{base}}$: temperature of the SiPMs during DCR measurement.
- $A_{s,\text{base}}$: active area of the SiPMs used during DCR measurement.
- f_{base} : fluence used to the irradiate SiPMs for DCR measurement.
- T_s : is the temperature of the HGCAL hence the SiPM at which it will be operated, which is -30° C .
- f : is the fluence the SiPM depending on its location will receive over its lifetime of operation in HGCAL.

5.4.2 Testbeam and SiPM Noise Inputs

CMS HGCAL collaboration conducted testbeam measurement in January 2020 on both cast and injection molded scintillator tiles wrapped in ESR with SiPM using Fermi National

Accelerator Laboratory (FNAL) 120 GeV testbeam facility. The scintillator tiles used in testbeam were of dimensions $30 \times 30 \text{ mm}^2$ square tiles, and SiPM device used was Hamamatsu S13360–1350CS ($1.3 \times 1.3 \text{ mm}^2$) [63, 69].

The scintillator tiles response measured from testbeam are 35 p.e. for cast scintillator tiles, and 25 p.e. for injection molded with SiPM operated at voltage of 54.26 V, which is V_0 of 2.5V (I-V method) (equivalent to 3.0V when measured with gain method). Currently SiPM device class expected to used in HGCAL is Hamamatsu S14160 with $15 \mu\text{m}$ pixel size (dubbed as HDR15) and $2, 4, 9 \text{ mm}^2$ in area operated at V_0 of 2V (I-V method), using ratio of PDEs of these devices we can calculate PDE scale factor as,

$$= \frac{\text{PDE of S14160 at } V_O = 2V}{\text{PDE of S13360 at } V_O = 3V} = \frac{34.9}{40} = 0.8725 \quad (5.4)$$

this gives, MIP* value to be 30.5 p.e. for cast and 21.8 p.e. for injection molded scintillator tiles.

DCR measurement for HDR15 (2 mm^2) SiPMs irradiated to $5 \times 10^{13} \text{ n/cm}^2$ operated at $V_0 = 2V$ (I-V method) and at temperature -30° C is equivalent to RMS value of 19 p.e. with 15 ns integration time period [70].

Using the testbeam measurement of scintillator tile response and irradiated SiPM DCR measurements end-of-life scenario estimation of detector performance was done for combinations of scintillator types and HDR15 SiPMs areas.

5.4.3 Scenarios

Five combinations of scintillator material and SiPM active area were considered and for each combination the detector performance was evaluated. If there are multiple options

passing $S/N > 3$ requirement for the same tileboard, the option with the higher preference is selected. Two different scenarios were considered with preference order as:

- **Scenario A:**

1. Injection Molded Scintillator Tiles and SiPM of active area 2 mm^2 .
2. Injection Molded Scintillator Tiles and SiPM of active area 4 mm^2 .
3. Cast Scintillator Tiles and SiPM of active area 2 mm^2 .
4. Cast Scintillator Tiles and SiPM of active area 4 mm^2 .
5. Cast Scintillator Tiles and SiPM of active area 9 mm^2 .

- **Scenario B:**

1. Injection Molded Scintillator Tiles and SiPM of active area 2 mm^2 .
2. Cast Scintillator Tiles and SiPM of active area 2 mm^2 .
3. Injection Molded Scintillator Tiles and SiPM of active area 4 mm^2 .
4. Cast Scintillator Tiles and SiPM of active area 4 mm^2 .
5. Cast Scintillator Tiles and SiPM of active area 9 mm^2 .

S/N of each combination after 3000 fb^{-1} when used alone is shown in Figure 5.11. Injection molded scintillator can be used in the deeper layers, and the cast scintillator with 9 mm^2 SiPM will be required for starting layers.

5.5 Results and Conclusion

For assembly of scintillator tiles on a tileboard, it is preferred to have single type of scintillator with SiPM combination. For this reason, each scenario is evaluated in the preference

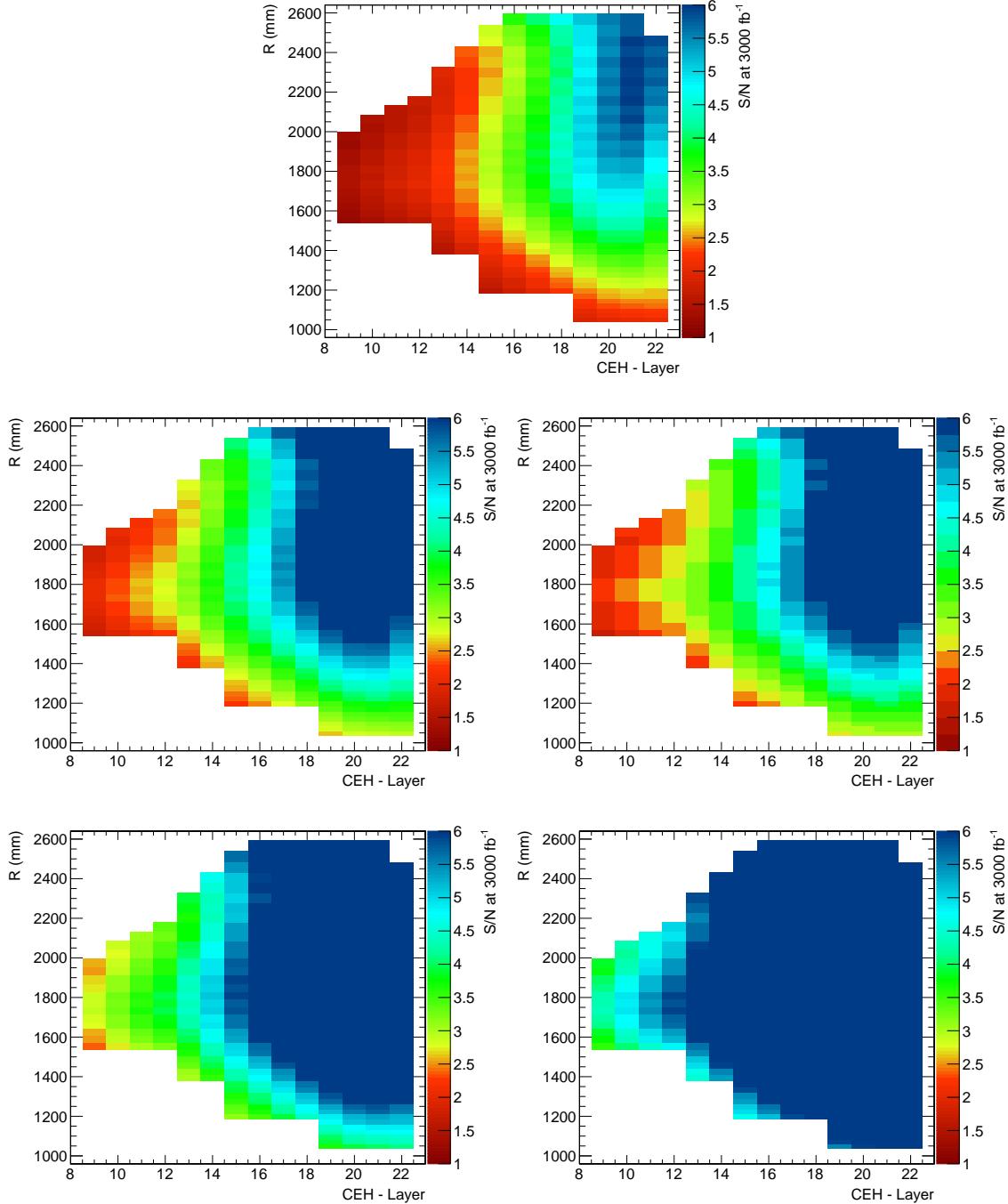


Figure 5.11: Scintillator performance with various active area size of SiPM. Top row from Left to Right: Injection Molded Scintillator with SiPM 2 and 4 mm^2 active area device. Bottom row from Left to Right: Cast Scintillator with SiPM 2, 4 and 9 mm^2 active area device.

order and tileboard is assigned a combination only if all the rings in it are able to satisfy $S/N > 3$. Figure 5.12 shows final results of how both scenarios fill tileboards in CE-H.

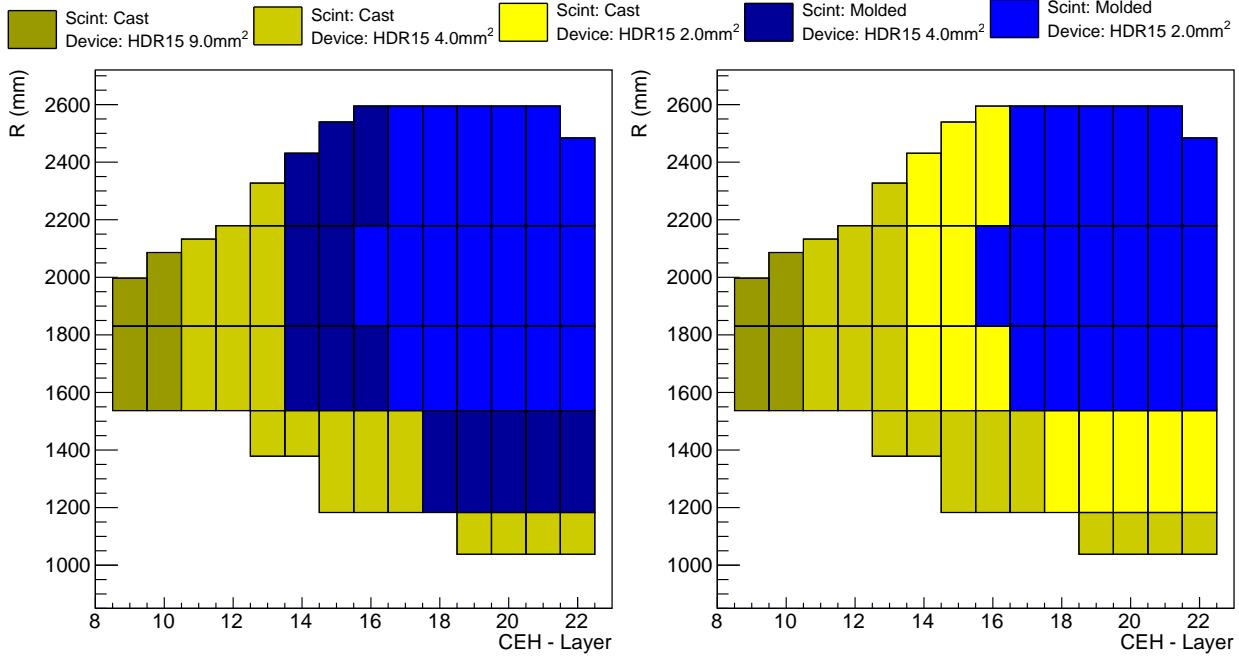


Figure 5.12: HGCAL scenarios. Left: Scenario A, Right: Scenario B

Table 5.1: HGCAL scenarios comparison

		Scenario A	Scenario B
Cast Scintillator	Cell Count	84, 096	153,216
	Total Area	101.59 m ²	197.46 m ²
	Percentage	27.8 %	54.0 %
Injection Molded Scintillator	Cell Count	155, 520	86, 400
	Total Area	264.04 m ²	168.17 m ²
	Percentage	72.2 %	46.0 %
SiPMs Count	2 mm ²	86, 400	155, 520
	4 mm ²	138, 240	69, 120
	9 mm ²	14, 976	14, 976

Both the scenarios require the same number of 9 mm² SiPMs (Table 5.1). Scenario B uses approximately half the 4 mm² SiPMs compared to scenario A. The amount of cast

and injection molded scintillator is approximately equal in scenario B. Scenario A has the advantage since it will require a smaller number of cast scintillator tiles to be produced. However, subsequent measurement of injection molded tiles from different production batches have shown that its light response can fluctuate by 20%. To be on the safe side the HGCAL collaboration has chosen scenario B for the scintillator section of HGCAL.

CHAPTER 6

SUMMARY

In this dissertation the related contribution of the development of SM VBS analysis in semileptonic $ZVjj$ channel, and the development and instrumentation of the upgraded endcap calorimeter (HGCAL) for the CMS experiment are discussed. Measurement of VBS is key to understanding the nature of EWSB at the TeV scale. Higher integrated luminosity and the HGCAL upgrade will improve the signal sensitivities of VBS like searches in the future.

This SM VBS analysis is done with 137 fb^{-1} of using 13 TeV proton-proton collision data collected by the CMS experiment during the 2016 to 2018 run period. The Z boson decays to a pair of leptons (electrons or muons), and the other vector boson V (either W or Z) decays hadronically to a pair of jets, in association of two large pseudorapidity jets. In this channel DY plus jets is the dominant background. To correct and normalize DY plus jets background model, a control region defined using hadronic boson mass was used. MVA approach was used to model the signal versus background classifier using gradient boosted BDT in the signal region. Expected significance of 1.5σ is reported for EW VBS $ZVjj$.

The HGCAL in addition to replacing the dying ECAL and HCAL hardware, will also enhance many VBS analyses. Since the jets in endcap suffer the most from pileup contamination. Reconstructing narrow jets using the lateral and longitudinal granularity of the silicon and scintillator cells, can significantly increase the physis sensitivity of the CMS data. HGCAL is expected to be installed during LS3 which is currently scheduled to be from end of year 2025 to the start of year 2029. Simulation and instrumentation studies, critical to the configuration and assembly of the scintillator part of the HGCAL are reported on.

REFERENCES

- [1] Wikipedia contributors. “Standard Model of Elementary Particles” (July 2021). Accessed September 07, 2021. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- [2] C. N. Yang and R. L. Mills. “Conservation of Isotopic Spin and Isotopic Gauge Invariance”. *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. DOI: [10.1103/PhysRev.96.191](https://doi.org/10.1103/PhysRev.96.191). URL: <https://link.aps.org/doi/10.1103/PhysRev.96.191>.
- [3] S. L. Glashow. “The renormalizability of vector meson interactions”. *Nuclear Physics* 10 (Feb. 1959), pp. 107–117. DOI: [10.1016/0029-5582\(59\)90196-8](https://doi.org/10.1016/0029-5582(59)90196-8).
- [4] S. Weinberg. “A Model of Leptons”. *Physical Review Letters* 19 (Nov. 1967), pp. 1264–1266. DOI: [10.1103/physrevlett.19.1264](https://doi.org/10.1103/physrevlett.19.1264).
- [5] A. Salam and J. C. Ward. “Weak and electromagnetic interactions”. *Il Nuovo Cimento* 11 (Feb. 1959), pp. 568–577. DOI: [10.1007/bf02726525](https://doi.org/10.1007/bf02726525).
- [6] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. *Physical Review Letters* 13 (Aug. 1964), pp. 321–323. DOI: [10.1103/physrevlett.13.321](https://doi.org/10.1103/physrevlett.13.321).
- [7] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. *Physical Review Letters* 13 (Oct. 1964), pp. 508–509. DOI: [10.1103/physrevlett.13.508](https://doi.org/10.1103/physrevlett.13.508).
- [8] J. Ellis. “Higgs Physics”. en (2015). DOI: [10.5170/CERN-2015-004.117](https://doi.org/10.5170/CERN-2015-004.117).
- [9] B. W. Lee, C. Quigg, and H. B. Thacker. “Weak interactions at very high energies: The role of the Higgs-boson mass”. *Physical Review D* 16 (Sept. 1977), pp. 1519–1531. DOI: [10.1103/physrevd.16.1519](https://doi.org/10.1103/physrevd.16.1519).

- [10] B. W. Lee, C. Quigg, and H. B. Thacker. “Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass”. *Physical Review Letters* 38 (Apr. 1977), pp. 883–885. DOI: 10.1103/physrevlett.38.883.
- [11] A. Denner and T. Hahn. “Radiative Corrections to $W^+W^- \rightarrow W^+W^-$ in the Electroweak Standard Model”. *Nucl.Phys.B* 525:27-50, 1998 (Nov. 12, 1997). DOI: 10.1016/S0550-3213(98)00287-9.
- [12] CMS Collaboration. “Evidence for WW/WZ vector boson scattering in the decay channel $l\nu qq$ produced in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV” (Dec. 10, 2021). arXiv: 2112.05259 [hep-ex].
- [13] CMS Collaboration. “Search for anomalous electroweak production of vector boson pairs in association with two jets in proton-proton collisions at 13 TeV”. *Phys. Lett. B* 798 (2019) 134985 (May 17, 2019). DOI: 10.1016/j.physletb.2019.134985. arXiv: 1905.07445 [hep-ex].
- [14] L. Evans and P. Bryant. “LHC Machine”. *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.
- [15] E. Mobs. “The CERN accelerator complex - 2019” (July 2019). Accessed August 16, 2021. URL: <http://cds.cern.ch/record/2684277/files/>.
- [16] Public CMS Luminosity Information. “Cumulative delivered and recorded luminosity versus time for run-2 pp data” (2018). Accessed August 17, 2021. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [17] The CMS Collaboration. “CMS Luminosity Measurements for the 2016 Data Taking Period” (2017). URL: <https://cds.cern.ch/record/2257069>.
- [18] The CMS Collaboration. “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV” (2018). URL: <https://cds.cern.ch/record/2621960>.

- [19] The CMS Collaboration. “CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV” (2019). URL: <https://cds.cern.ch/record/2676164>.
- [20] T. Sakuma. “Cutaway diagrams of CMS detector” (May 2019). Accessed August 17, 2021. URL: <https://cds.cern.ch/record/2665537>.
- [21] The CMS Collaboration. “The CMS experiment at the CERN LHC”. *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: [10.1088/1748-0221/3/08/s08004](https://doi.org/10.1088/1748-0221/3/08/s08004). URL: <https://doi.org/10.1088/1748-0221/3/08/s08004>.
- [22] The CMS Collaboration. “CMS Detector Slice” (Jan. 2016). Accessed August 18, 2021. URL: <https://cds.cern.ch/record/2120661>.
- [23] M. Brice and C. Marcelloni. “The central part of CMS is lowered” (Feb. 2007). Accessed August 22, 2021. URL: <https://cds.cern.ch/record/1020311>.
- [24] T. Sakuma. “SketchUp Drawings of the Phase 1 pixel detector alongside the current pixel detector” (June 2012). Accessed August 17, 2021. URL: <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB>ShowDocument?docid=6473>.
- [25] The CMS Collaboration. “The CMS ECAL performance with examples”. *Journal of Instrumentation* 9.02 (Feb. 2014), pp. C02008–C02008. DOI: [10.1088/1748-0221/9/02/c02008](https://doi.org/10.1088/1748-0221/9/02/c02008).
- [26] The CMS Collaboration. “Images of CMS ECAL Endcap (EE)” (Nov. 2008). Accessed August 22, 2021. URL: <https://cds.cern.ch/record/1431479>.
- [27] CMS Collaboration. “Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\text{sqrt}(s) = 7$ TeV”. *JINST* 8 (2013) P09009 (June 9, 2013). DOI: [10.1088/1748-0221/8/09/P09009](https://doi.org/10.1088/1748-0221/8/09/P09009). arXiv: 1306.2016 [hep-ex].
- [28] J. Mans et al. “CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter” (Sept. 2012). URL: <https://cds.cern.ch/record/1481837>.

- [29] The CMS Collaboration. “Images of the CMS HCAL Barrel (HB)” (2008). Accessed August 22, 2021. URL: <https://cds.cern.ch/record/1431485>.
- [30] The CMS Collaboration. “HCAL Depth Segmentation Phase 1 Upgrade” (2012). Accessed August 22, 2021. URL: <https://home.fnal.gov/~chlebana/CMS/Phase1>.
- [31] The CMS Collaboration. “Performance of the CMS Muon Detectors in early 2016 collision runs” (2016). Accessed August 22, 2021. URL: <https://twiki.cern.ch/twiki/pub/CMSPublic/MuonDPGPublic160729>.
- [32] The CMS Collaboration. “Particle-flow reconstruction and global event description with the CMS detector”. *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–P10003. DOI: [10.1088/1748-0221/12/10/p10003](https://doi.org/10.1088/1748-0221/12/10/p10003).
- [33] S. Cucciarelli et al. “Track reconstruction, primary vertex finding and seed generation with the Pixel Detector”. CMS-NOTE-2006-026 (Jan. 2006). URL: <https://cds.cern.ch/record/927384>.
- [34] W. Adam et al. “Track Reconstruction in the CMS tracker”. CMS-NOTE-2006-041 (Dec. 2006). URL: <https://cds.cern.ch/record/934067>.
- [35] CMS Collaboration. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JINST* 13 (2018) P06015 (Apr. 12, 2018). 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).
- [36] W. Adam et al. “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”. *Journal of Physics G: Nuclear and Particle Physics* 31.9 (July 2005), N9–N20. DOI: [10.1088/0954-3899/31/9/n01](https://doi.org/10.1088/0954-3899/31/9/n01).
- [37] The CMS Collaboration. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. *JINST* 16 (2021) P05014 (Dec. 12, 2020). DOI: [10.1088/1748-0221/16/05/P05014](https://doi.org/10.1088/1748-0221/16/05/P05014).

- [38] M. Cacciari, G. P. Salam, and G. Soyez. “FastJet user manual” (Nov. 25, 2011). DOI: 10.1140/epjc/s10052-012-1896-2.
- [39] D. Bertolini et al. “Pileup Per Particle Identification”. *JHEP* 1410 (2014) 59 (July 22, 2014). DOI: 10.1007/JHEP10(2014)059.
- [40] CMS Collaboration. “Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018” (Apr. 2020). URL: <https://cds.cern.ch/record/2715872>.
- [41] The CMS Collaboration. “Pileup mitigation at CMS in 13 TeV data”. *Journal of Instrumentation* 15.09 (Sept. 2020), P09018–P09018. DOI: 10.1088/1748-0221/15/09/p09018.
- [42] J. Thaler and K. V. Tilburg. “Identifying Boosted Objects with N-subjettiness”. *JHEP* 1103:015,2011 (Nov. 10, 2010). DOI: 10.1007/JHEP03(2011)015.
- [43] The CMS Collaboration. “Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques”. *JINST* 15 (2020) P06005 (Apr. 17, 2020). DOI: 10.1088/1748-0221/15/06/P06005.
- [44] A. J. Larkoski et al. “Soft Drop”. *JHEP* 1405 (2014) 146 (Feb. 11, 2014). DOI: 10.1007/JHEP05(2014)146.
- [45] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations” (May 1, 2014). DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].
- [46] P. Artoisenet et al. “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”. *Journal of High Energy Physics, Volume 2013, Issue 3* (Dec. 14, 2012). DOI: 10.1007/JHEP03(2013)015. arXiv: 1212.3460 [hep-ph].

- [47] B. Andersson et al. “Parton fragmentation and string dynamics”. *Physics Reports* 97.2-3 (July 1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.
- [48] CMS Collaboration. “Event generator tunes obtained from underlying event and multiparton scattering measurements”. *Eur. Phys. J. C* 76 (2016) 155 (Dec. 2, 2015). DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].
- [49] CMS Collaboration. “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”. *Eur. Phys. J. C* 80 (2020) 4 (Mar. 28, 2019). DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].
- [50] C. Oleari. “The POWHEG-BOX”. *Nucl.Phys.Proc.Suppl.* 205-206:36-41,2010 (July 22, 2010). DOI: 10.1016/j.nuclphysbps.2010.08.016. arXiv: 1007.3893 [hep-ph].
- [51] CMS Collaboration. “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JINST* 15 (2020) P10017 (June 17, 2020). DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].
- [52] A. Hoecker et al. “TMVA - Toolkit for Multivariate Data Analysis” (Mar. 4, 2007). arXiv: physics/0703039 [physics.data-an].
- [53] D. Rainwater, R. Szalapski, and D. Zeppenfeld. “Probing color-singlet exchange in $Z + 2$ -jet events at the LHC”. *Phys.Rev.D* 54:6680-6689,1996 (May 30, 1996). DOI: 10.1103/PhysRevD.54.6680. arXiv: hep-ph/9605444 [hep-ph].
- [54] W. Verkerke and D. Kirkby. “The RooFit toolkit for data modeling” (June 14, 2003). arXiv: physics/0306116 [physics.data-an].
- [55] L. Moneta et al. “The RooStats Project” (Sept. 6, 2010). arXiv: 1009.1003.
- [56] R. Barlow and C. Beeston. “Fitting using finite Monte Carlo samples”. *Computer Physics Communications* 77.2 (Oct. 1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-w.

- [57] A. Ballestrero et al. “Precise predictions for same-sign W-boson scattering at the LHC”. *Eur.Phys.J. C78* (2018) no.8, 671 (Mar. 21, 2018). DOI: 10.1140/epjc/s10052-018-6136-y. arXiv: 1803.07943 [hep-ph].
- [58] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. *Eur.Phys.J.C71:1554,2011* (July 10, 2010). DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].
- [59] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group. “Procedure for the LHC Higgs boson search combination in Summer 2011”. CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11 (Aug. 2011). URL: <https://cds.cern.ch/record/1379837>.
- [60] CMS Collaboration. “The Phase-2 Upgrade of the CMS Endcap Calorimeter” (Nov. 2017). DOI: 10.17181/CERN.1V8M.1JY2. URL: <https://cds.cern.ch/record/2293646>.
- [61] D. Barney. “Overview slide of CE with main parameters” (Oct. 2019). Accessed August 24, 2021. URL: <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB>ShowDocument?docid=13251>.
- [62] D. Barney and T. Sakuma. “Sketchup images highlighting the sub-detectors” (Sept. 2017). Accessed August 24, 2021. URL: <https://cds.cern.ch/record/2628519>.
- [63] Hamamatsu. “S13360 series MPPC” (June 2022). Accessed June 13, 2022. URL: https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRARY/ssd/s13360_series_kapd1052e.pdf.
- [64] Hamamatsu. “S14160 series MPPC” (June 2022). Accessed June 13, 2022. URL: https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRARY/ssd/s14160_s14161_series_kapd1064e.pdf.

- [65] V. Zutshi et al. “Test Beam Performance of Directly Coupled Scintillator Tiles using Hamamatsu MPPCs”. *Physics Procedia* 37 (2012), pp. 789–795. DOI: [10.1016/j.phpro.2012.02.426](https://doi.org/10.1016/j.phpro.2012.02.426).
- [66] G. Blazey et al. “Directly coupled tiles as elements of a scintillator calorimeter with MPPC readout”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 605.3 (July 2009), pp. 277–281. DOI: [10.1016/j.nima.2009.03.253](https://doi.org/10.1016/j.nima.2009.03.253).
- [67] R. Singh. “Automated Scintillator Tile Wrapping at NIU” (Mar. 2022).
- [68] J. Mans and P. de Barbaro. “The baseline readout of the scintillator section of the CMS HL-LHC Endcap Calorimeter” (Feb. 2017).
- [69] A. Belloni et al. “Test beam study of SiPM-on-tile configurations”. *Journal of Instrumentation* 16.07 (July 2021), P07022. DOI: [10.1088/1748-0221/16/07/p07022](https://doi.org/10.1088/1748-0221/16/07/p07022).
- [70] Y. Musienko. “Recent measurements with SiPM prototypes” (June 2020). URL: https://indico.cern.ch/event/927798/contributions/3900920/attachments/2054713/3445624/Recent_measurements_with_SiPM_prototypes_HGCAL-SiPM_meeting_10.06.2020-corr.pdf.

APPENDIX A
ANALYSIS CODE

Analysis code used for the analysis is hosted on Github (<https://github.com>) platform.

- Custom NanoAOD production from MiniAOD to include missing PDF weights.

<https://github.com/singh-ramanpreet/VBS-customNanoAODProduction/>

- “NanoSkim”: Intermediate skimming step for the analysis phase space with minimal selection to save time, when run it again during analysis development.

<https://github.com/singh-ramanpreet/VVjjSemileptonic-NanoSkim>

- “Selection”: This repo contains the code for main event selection of this analysis, it also calculates and embed scale factors for various objects.

<https://github.com/singh-ramanpreet/VVjjSemileptonic-Selection>

- “Analysis”: This repo contains code MVA training, MVA inference and embedding, making Data/MC histograms, making datacards for the statistical analysis with “combine”.

<https://github.com/singh-ramanpreet/VVjjSemileptonic-Analysis>

APPENDIX B

ADDITIONAL FIGURES

B.1 Control Region DY+Jets Boosted ZV Channel

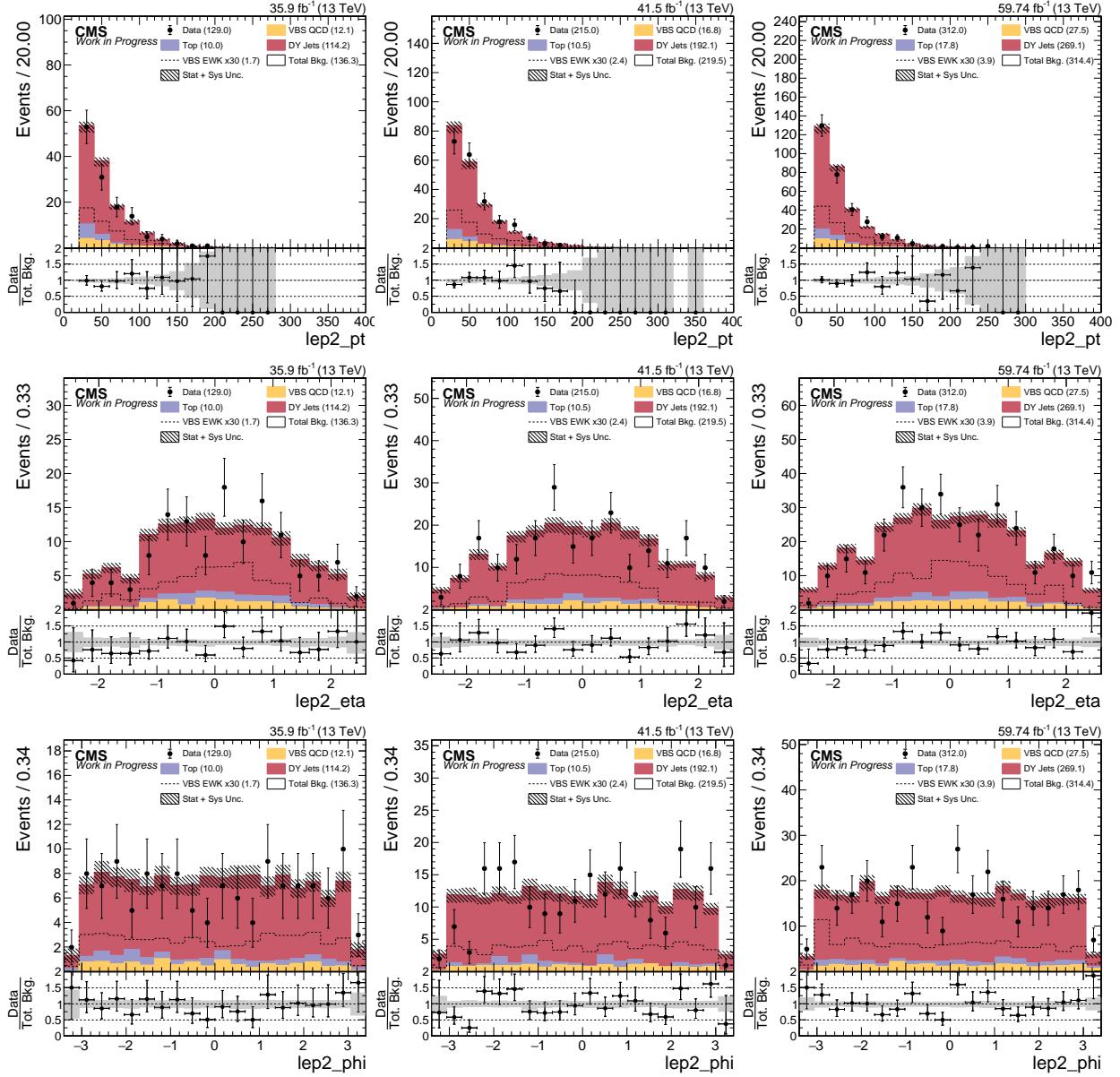


Figure B.1: DY+Jets Control Region: Trailing electron kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

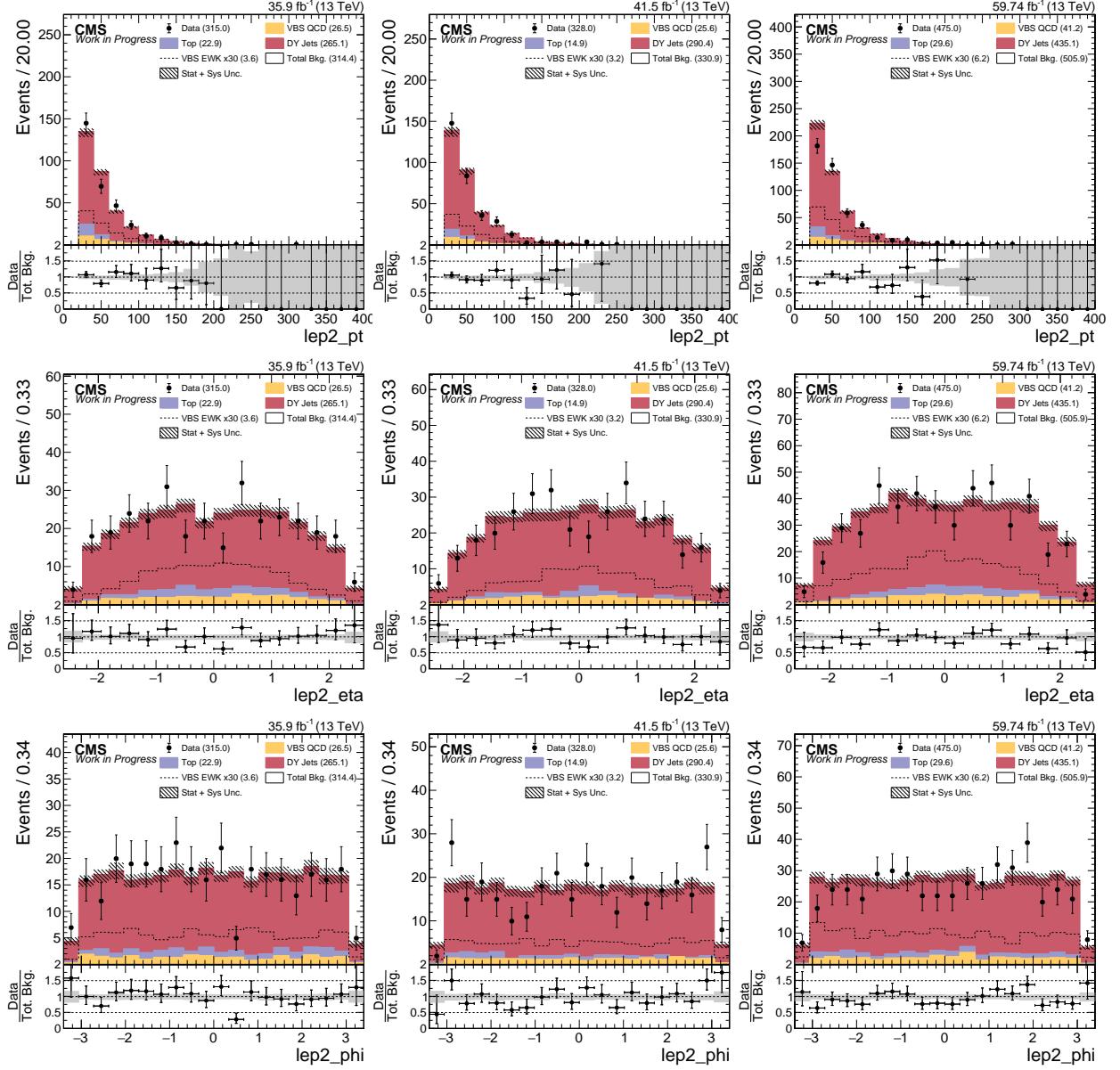


Figure B.2: DY+Jets Control Region: Trailing muon kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

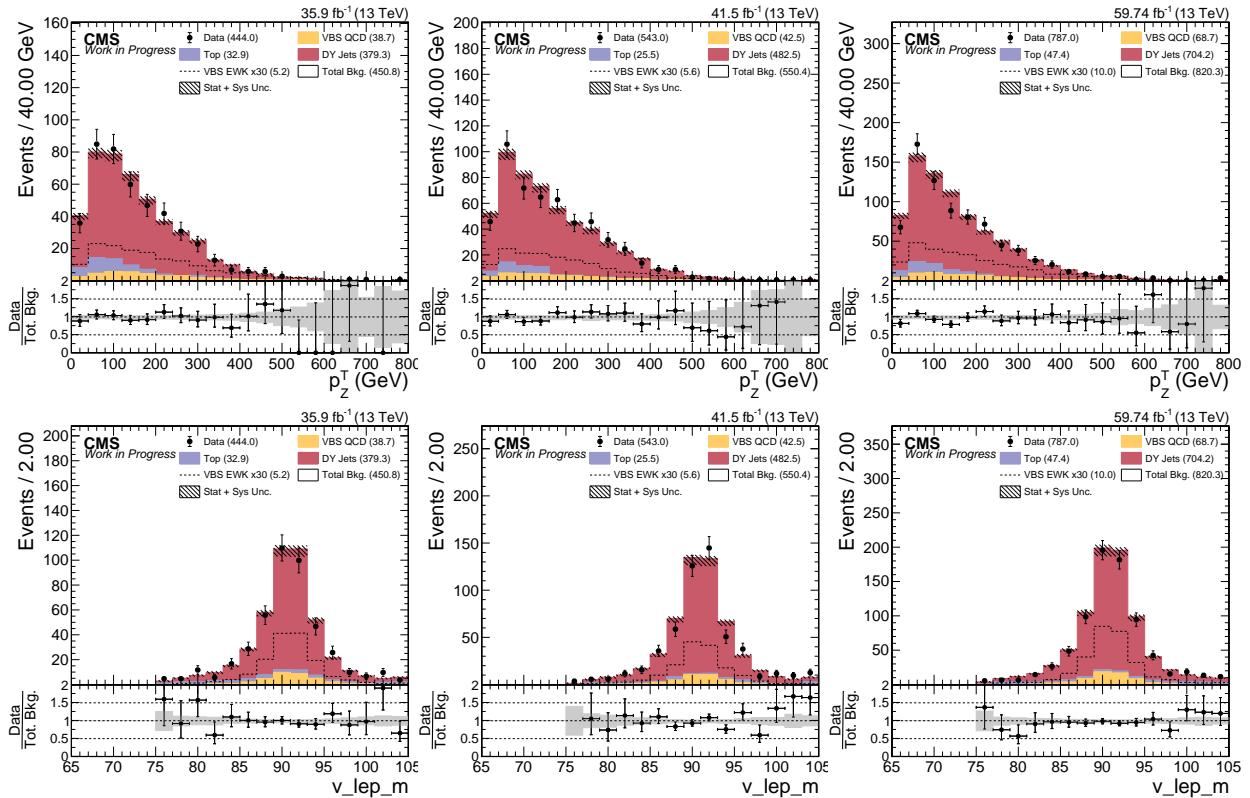


Figure B.3: DY+Jets Control Region: Z boson kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , and mass.

B.2 Control Region DY+Jets Resolved ZV Channel

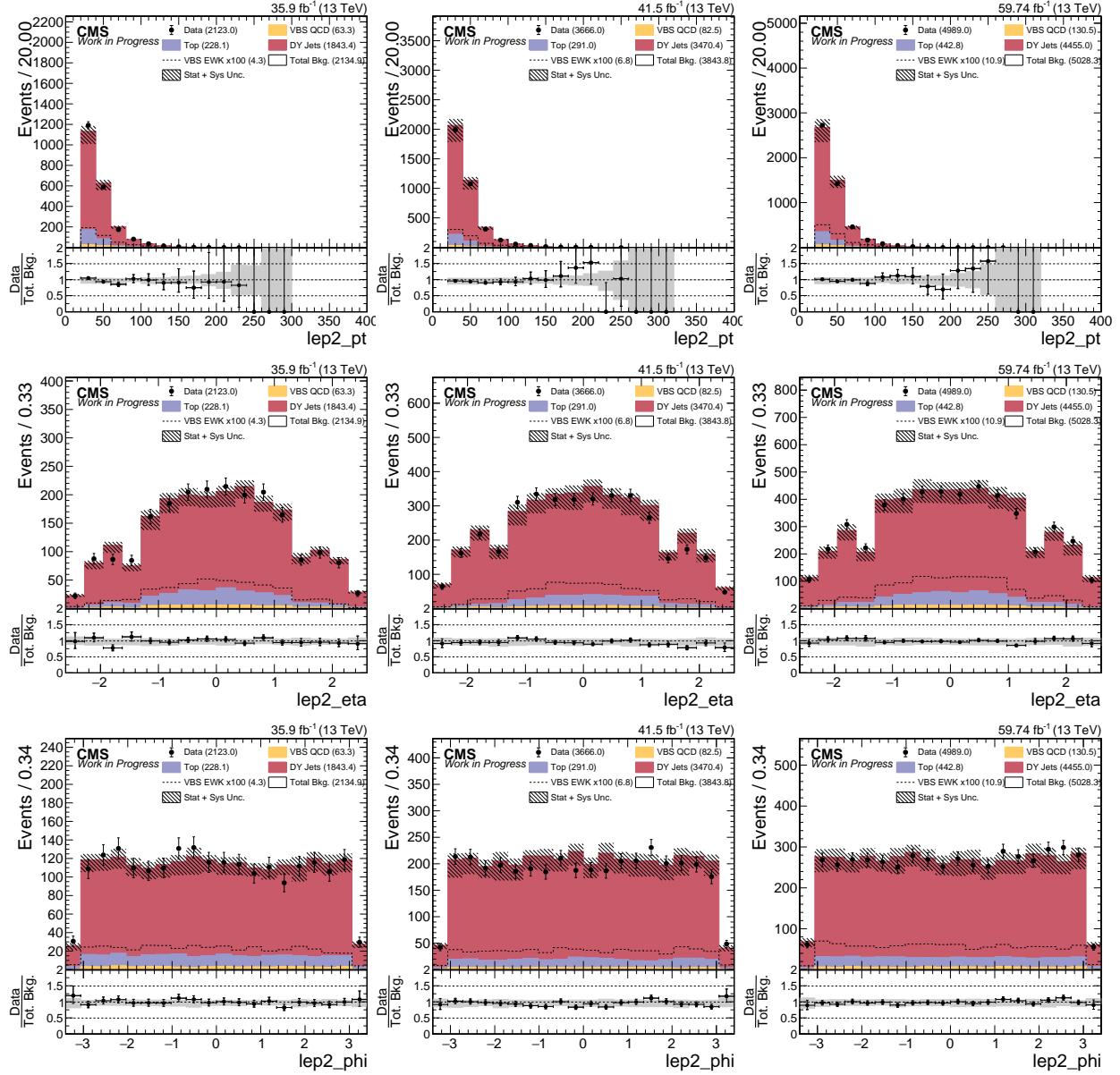


Figure B.4: DY+Jets Control Region: Trailing electron kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

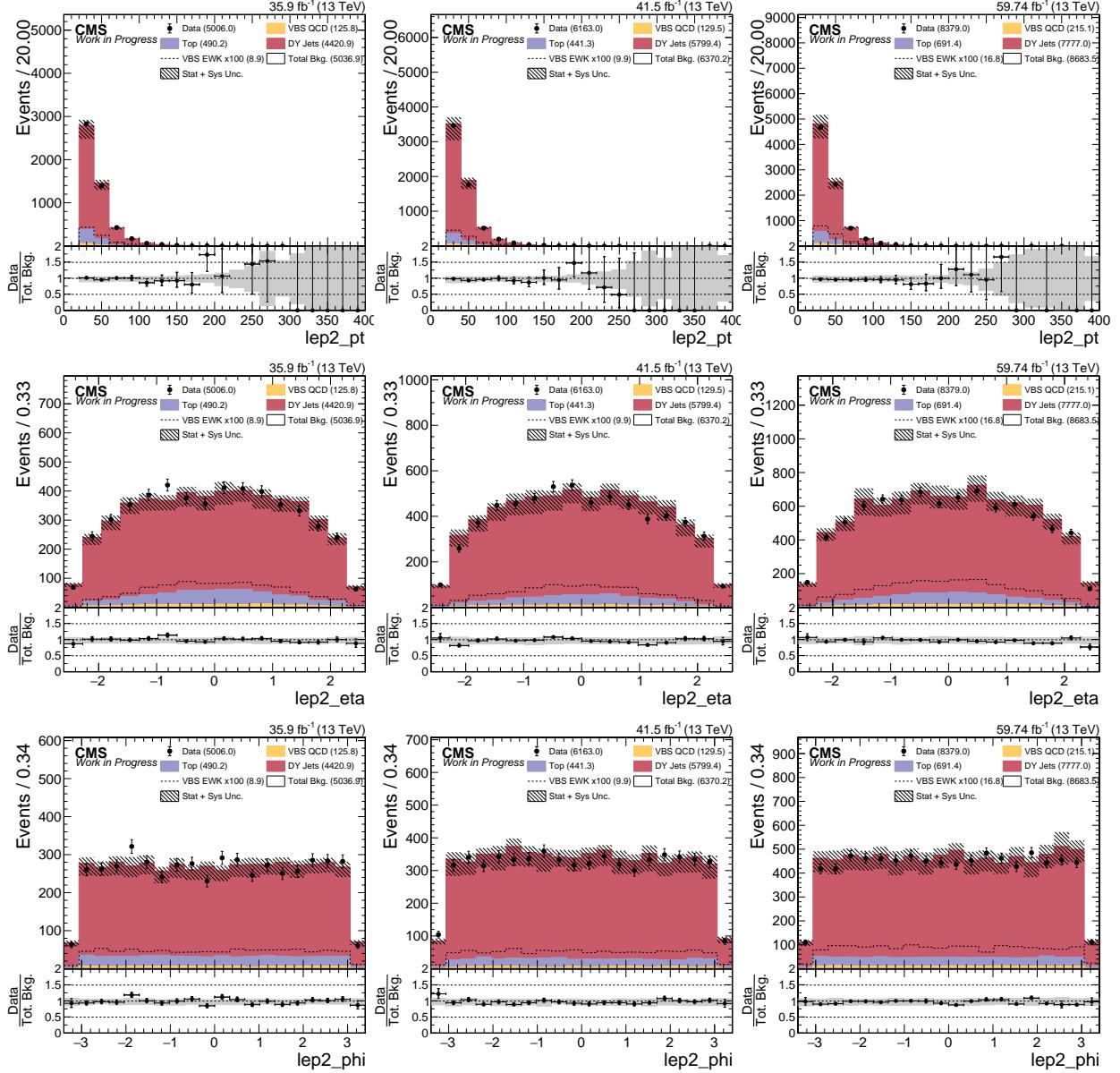


Figure B.5: DY+Jets Control Region: Trailing muon kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , and ϕ .

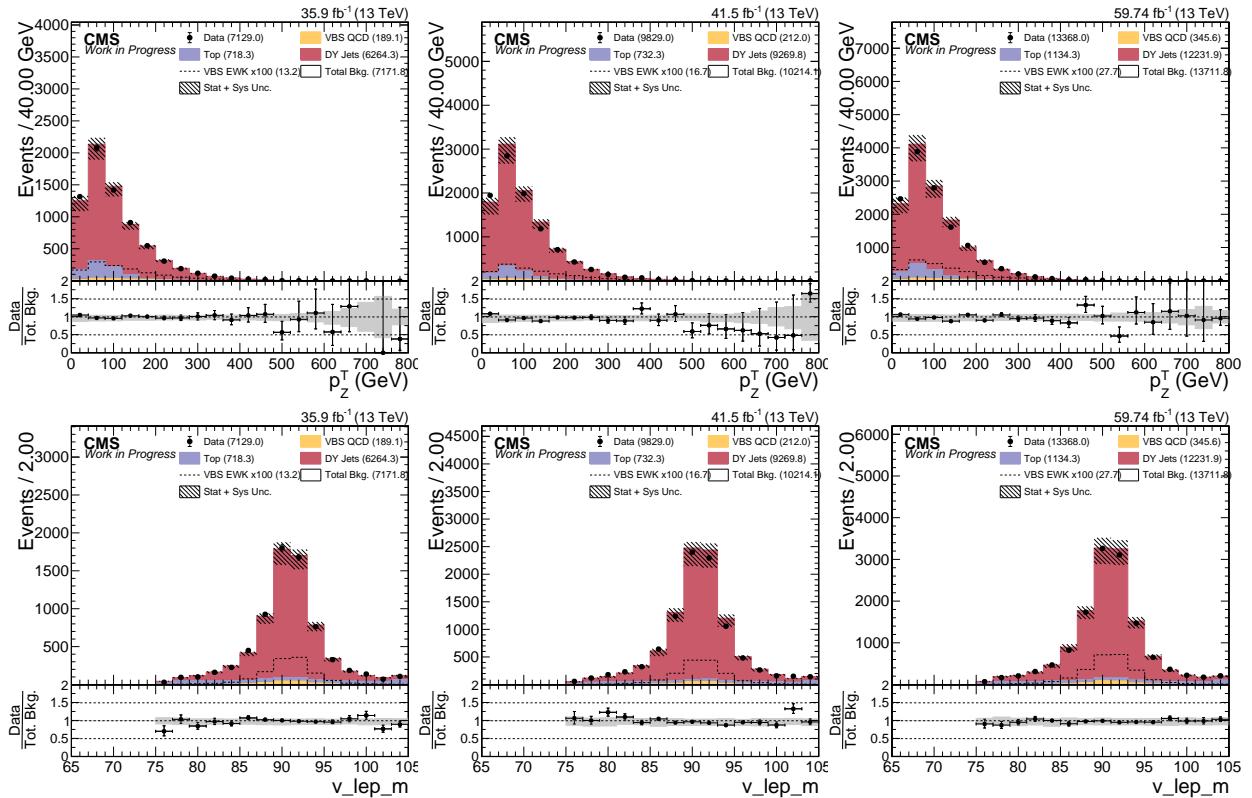


Figure B.6: DY+Jets Control Region: Z boson kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T^Z , and mass.

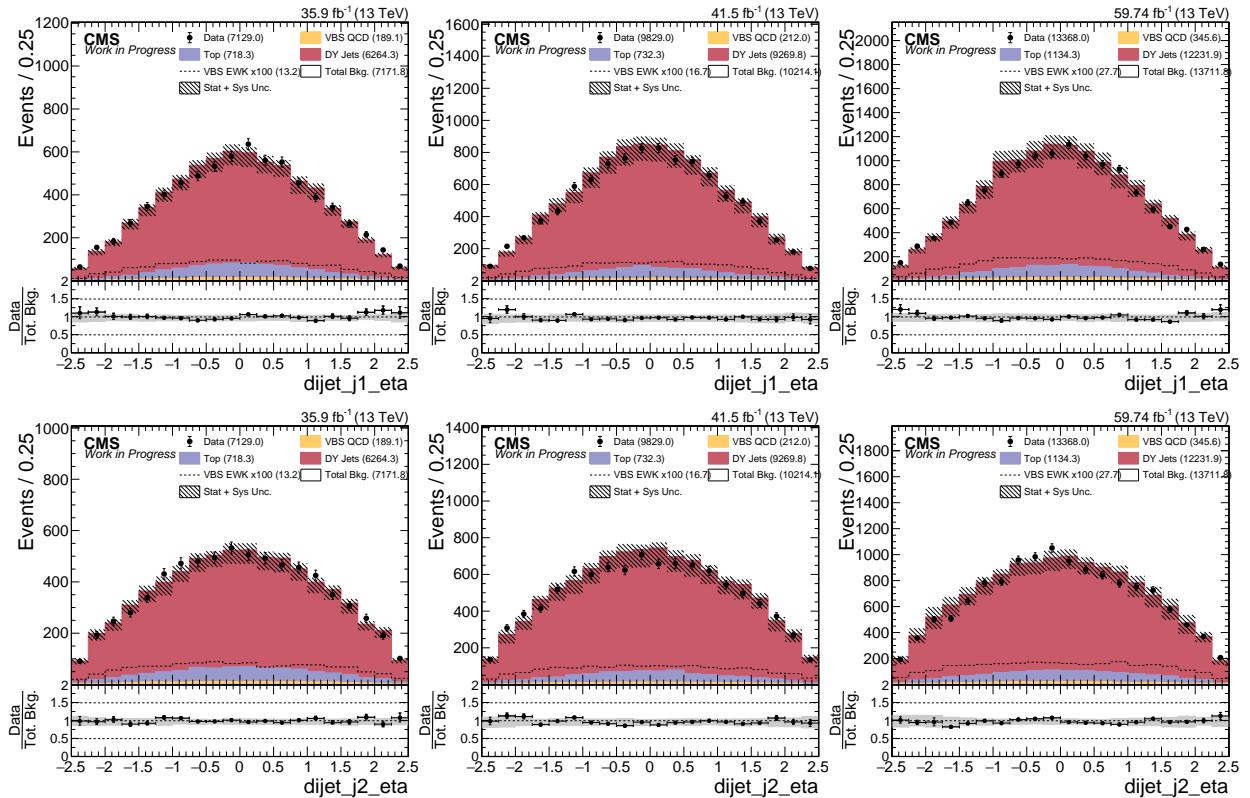


Figure B.7: DY+Jets Control Region: V boson leading and trailing jet η in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: η of leading jet, and η of trailing jet.

B.3 Impact Plots

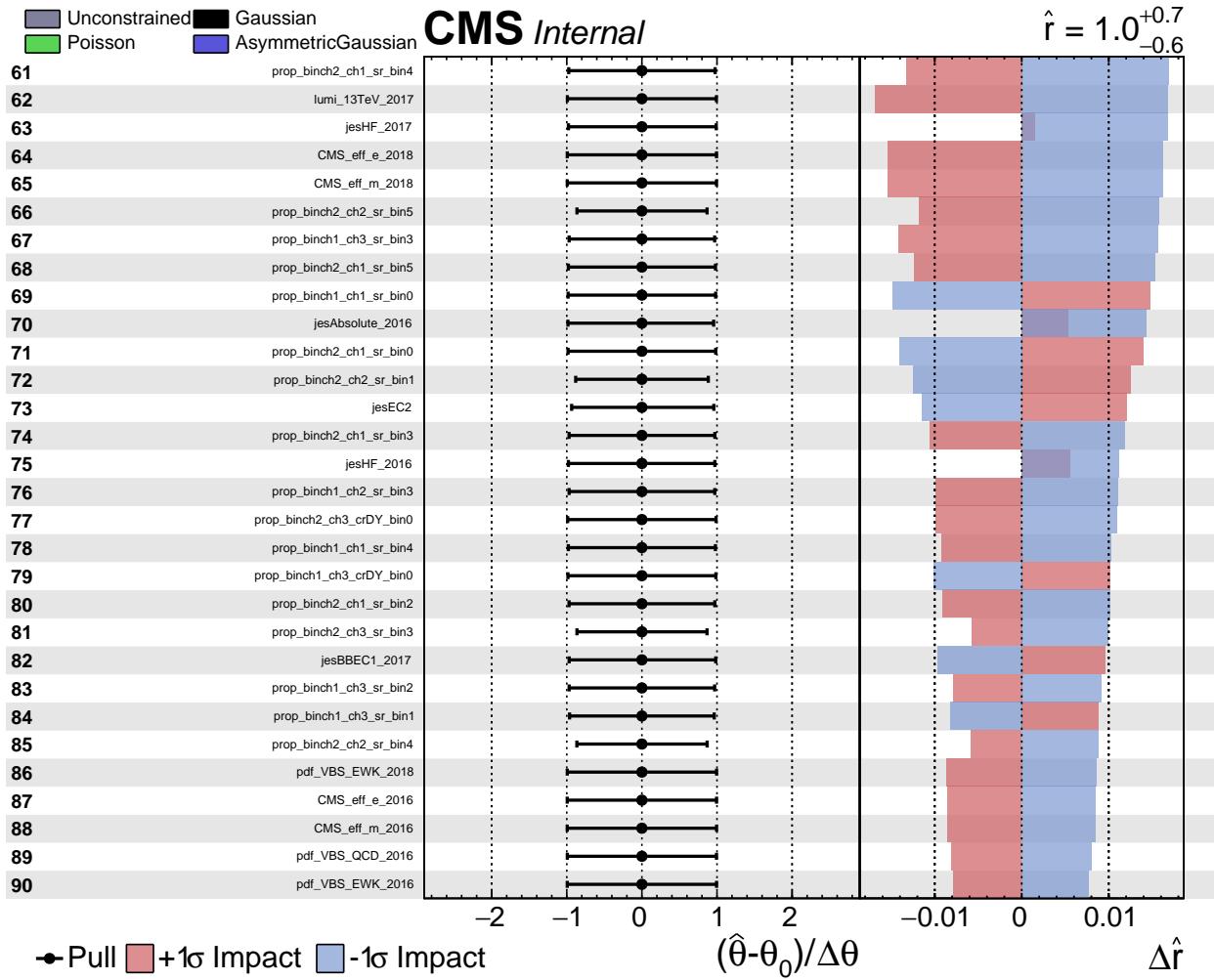


Figure B.8: Impact Plots of nuisance parameters from 61 to 90.

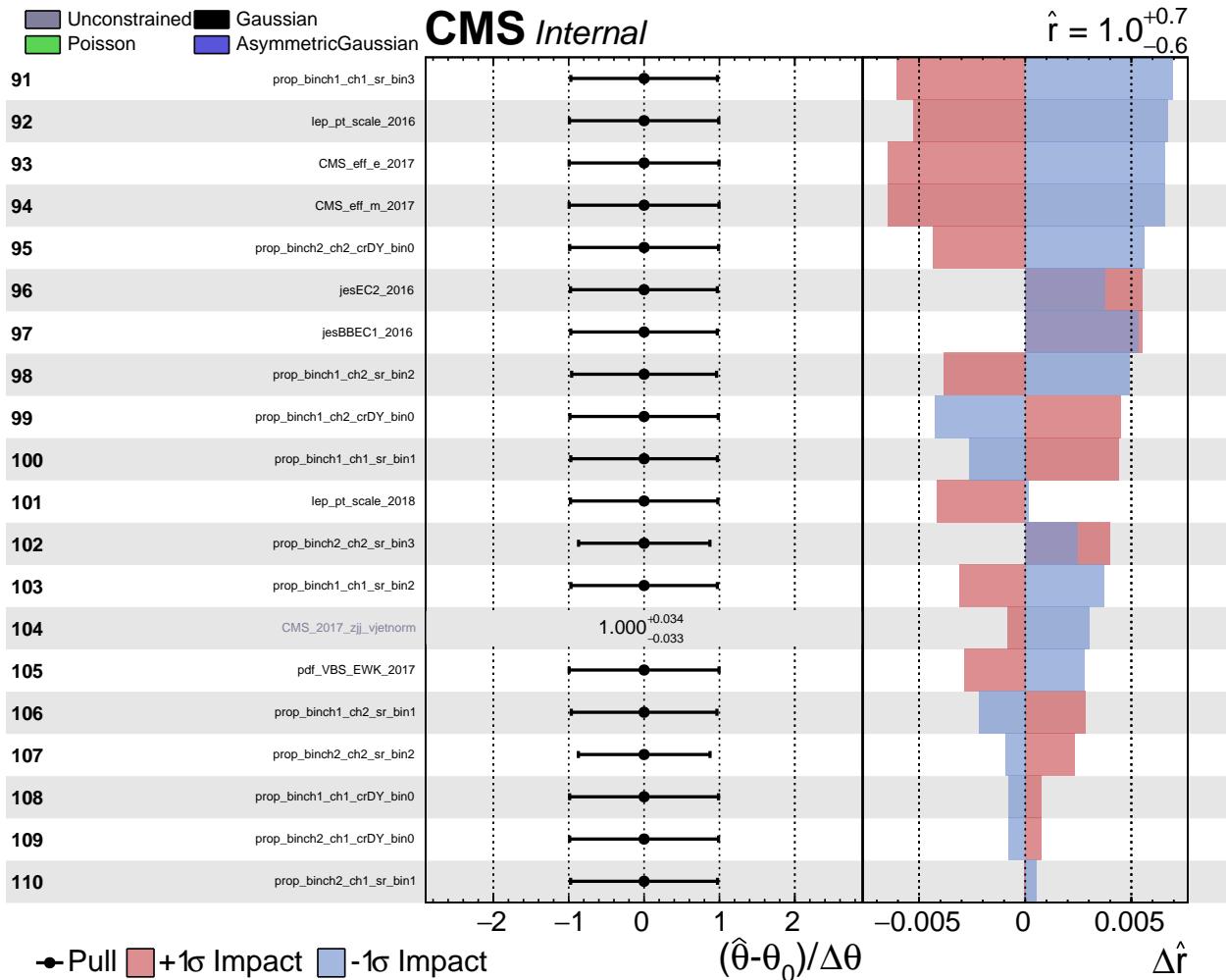


Figure B.9: Impact Plots of nuisance parameters from 91 to 110.